

Catechism for nuclear power

Governments should urgently give attention to the consequences of Chernobyl on public opinion of nuclear power — and should correct some of the damage they have done.

HERE is a question from an examination paper in the final examination for students following the course in Logic for the Modern World offered by the University of the Sticks, on which people may wish to brood

Q. "Even small doses of radiation exposure are potentially harmful to people, and should be avoided."

"Nuclear power stations, and the nuclear industry in general, are sources of artificial radiation to which people may be exposed."

"Therefore nuclear power stations should not be built."

Either explain in no more than 250 words which of the three statements in this syllogism is invalid **or** outline in no more than 50,000 words an alternative energy policy for your country of nationality drawing attention to such social consequences as you can foresee."

Like most examination questions, this example begs a host of other questions, mostly unexamined. But the example is not nearly as hypothetical as it may seem. Especially in the aftermath of the Chernobyl accident but not exclusively on that account the imputed logic is increasingly coming to dominate the policies of governments across the world. And while there remains what may be called a hard core of people who believe that nuclear power remains an economic necessity, the syllogism cannot as easily as they would wish be demolished in the prescribed 250 words.

Radiation

Those familiar with exercises like these will nevertheless quickly know the truth that the insidious errors are in the premise. Once that has been swallowed, the conclusion tends to follow as night follows day. Nuclear power stations and the nuclear industry in general are, indeed, sources of radiation to which people, either workers in the nuclear industry or the general population, may be exposed. The British population, and British farmers in particular, have recently been given a dramatic proof of that by the decision of the agriculture ministry to ban the sale of fat lambs on the flimsy grounds that some were found to contain just over 10 per cent of the amount of Chernobyl caesium that would have been permitted by the emergency reference levels promulgated earlier this year by the British National Radiological Protection Board (see *Nature* 321, 798; 1986). Earlier, the whole European Community had been reminded that power stations are potentially powerful sources of radiation by the decision, at Brussels, that imports of foodstuffs from Eastern Europe would be prohibited while the content of radioactivity exceeded 600 Bq per kg, 6 per cent of the same reference level. The effect of these and other decisions has been to dramatize the truth that nuclear power stations are potential sources of radioactivity, which may be geographically widespread. The lesson will now be firmly embedded in the public consciousness, as it should be. So, given the premise, the conclusion follows that nuclear power stations are an abomination and should not be built.

As it happens, the bulk of the premise is also correct. "Even small doses of radiation exposure" are indeed potentially harmful to people. The days have long since gone when it was permissible for politicians and others to canvass the idea that there is a

"threshold" dose below which no damage of any kind is done to living organisms. While much remains to be learned about the biological consequences of small doses of radiation, the best assumption is that the probability of a particular kind of damage is proportional to the radiation dose, however small. To be sure, it is possible that in the causation of some kinds of cancer by radiation (or even other agents), where two separate damaging interactions between radiation and tissues may be required to trigger the neoplastic process, the risk of damage may be proportional to the square of the dose. But, in the present state of ignorance, nobody would go to the stake for that hypothesis, which would in any case only partially restore the respectability of the notion that there are thresholds below which no damage of this particular kind — carcinogenesis — is done. And where genetic damage is in question, there seems (at least for the time being) no sensible alternative to the proposition that the risk of damage is proportional to the dose.

Risk

The circumstance that different kinds of radiation exposure are potentially damaging in different ways does not undermine the validity of the first part of the premise in the examination question, although this circumstance enormously complicates both the estimation of risk and the general understanding of why radiation is hazardous. It is now a familiar circumstance that different sources of radiation, natural or artificial, must be dealt with differently in the estimation of biological effects. Naturally occurring radon (on account of its decay products) and airborne radioactivity are potentially causes of lung cancer to the extent that they are trapped in the airways of the lungs. Naturally occurring potassium-40 and artificial caesium (as in British post-Chernobyl lambs) once ingested, expose the internal organs of the body almost equally to potentially damaging doses of radiation. Both plutonium-239 and strontium-90 concentrate in bones and are a threat to bone marrow, but, becquerel for becquerel, plutonium-239 is more damaging than strontium because alpha particles deposit their energy along shorter paths of ionization. And so on. There is far too little public understanding of the sophistication now applied by the radiological protection people as a matter of routine to the unravelling of the once-daunting complexity. Yet none of this qualifies the general principle that the potential hazard of a radiation dose (of a particular kind) is probably proportional to its amount.

The catch in the premise is the hanging qualifier in the premise, the seemingly innocent phrase "and should be avoided". This, of course, is not so much the starting point for a syllogism but the conclusion of some other. Moreover, whether premise or conclusion, it cannot be taken by reasonable people as an absolute injunction against radiation exposure. If it were such, the whole pattern of civilized life would already, during the half-century in which understanding of the hazards of radiation have accumulated, have been transformed. People would no longer live, or be allowed to live, at high altitudes or even high latitudes, on account of the extra doses of cosmic radiation to which they are there exposed. Old granite shield areas, where radon in the atmosphere accounts for up to a third of natural exposure,



would be evacuated. Jet aircraft would be disallowed. And so on.

The truth, of course, is that the principle that radiation exposure is best avoided is not now, and never can be, absolute. Merely to rationalize present patterns of life, people have to acknowledge that the heavily loaded part of the premise must be extended somehow, perhaps to read "should be avoided whenever possible" or, more tangibly, "should be weighed in a cost-benefit calculation and then, when the costs exceed the benefits, avoided". The snag for the simplifiers is that any such qualification undermines the logic of the syllogism. Only if the interdiction of radiation exposure is absolute does the conclusion follow that nuclear power stations must never be built. In all other circumstances, the most that can be said is that the benefits of nuclear power must be weighed against the costs, one of which is the potential damage done by radiation exposure, both as a matter of routine (to workers at reprocessing plants) and after accidents (as at Chernobyl).

All this is familiar stuff. For thirty years, those who build (or would build) nuclear power stations have been saying just this. It may be true that, for much of that period, the professionals have been less than frank about the potential scale of the hazards, but nobody has sought to conceal the need for a rational trade-off between risks and benefits. What Chernobyl might rationally have accomplished is a demonstration that, within a well-run nuclear programme, the avoidance of major reactor accidents deserves even more attention than it has received since the close attention given to the problem by the monumental Rasmussen study just over a decade ago. It is also possible to argue that it is worth paying something extra to avoid self-imposed risks, such as those unavoidably attending a nuclear industry, on the principle that there is a difference between unavoidable sources of risk (cosmic rays, for example) and those that are self-imposed. Yet even that seductive argument leaks. By what tests is it virtuous to claim (as the British Labour Party now does) that nuclear power should be replaced by coal when one of the few certainties in the trade-off is that many more miners will be killed per gigawatt-hour in the coalmining industry than members of the public by exposure to the radiation accompanying the routing operation of a nuclear industry?

If Chernobyl and its consequences were to lead to a more public re-examination of these questions in the countries where nuclear power is potentially an economic benefit, nobody would complain. The trouble is that the accident has come at a time when the populations most likely to need nuclear power in the decades ahead have been dissuaded from regarding the issues calmly by siren voices seeming to proclaim that there is such a thing as a free lunch — electricity without the risk to those who consume it. The sad truth is that governments are often unwittingly the abettors of these seductive propagandists, as when European governments band together to settle on a limit for the contamination of imported foodstuffs that provocatively plays over-safe. They seem not to have appreciated that, by taking such a line, they have gone a long way to accepting the loaded premise of the false syllogism which will be used against them when they plan to build a nuclear reactor. Is it too late to ask that they should mend their ways? □

Teaching by numbers

A British committee has produced an enlightening report on future trends in education.

WHAT is the difference between a teacher and a teaching machine? This question, widely asked in the sense pejorative of teachers a decade or more ago, has usually been answered by the claim (usually on behalf of teachers) that even the best teaching machines are merely mechanical, incapable of firing the imagination and the aspirations of students. For the most part, the teachers' case has been vindicated by the appalling quality of

what has previously been passed off under the label of educational technology. Some teaching machines have been literally machines whose mastery required not merely skills available to, say, high-school students, but those of ambidextrous conjurers as well. Others have looked like books but have been seen, on casual inspection, to be ways of drilling students in lessons which it is possible, but inappropriate, to learn by rote, leaving the real work to be done by real teachers. It is no wonder, that after a brief fashion for educational technology in the 1960s, the exercise should have been discredited. But equally, now that professional people are agog with the idea that "expert" systems should be used as aids to judgement in fields as different as medicine and engineering, it is natural that the old claims of educational technology should be dusted off and re-examined.

One result, in Britain, is a sensible slim document by a government committee called the Information Technology Advisory Panel, set up when Mr Kenneth Baker, now Secretary of State for Education and Science, was the British government's cheer leader for the information revolution. One irony is that most of the panel's recommendations urge that the government should spend money on investigation and research on the application of information technology; Mr Baker, in his new role, would have to foot the bill. Another is that the panel was subsumed last April in another government committee, the Advisory Committee on Applied Research and Development, whose most recent public report some weeks ago consisted of an extended wringing of hands over the parlous condition of the British software industry, from which many of the defunct panel's members spring. Yet, curiously, the panel's report (*Learning to live with IT*, HMSO £4.00) is just the judicious blend of enthusiasm for change and caution about the means by which it may be accomplished that endears a committee to its discriminating followers.

On the leading question, not directly answered by the panel, the difference between a teacher and a machine is that the teacher alone can serve as an intelligent critic of an individual student's learning. Machines, of course, can automatically assess a student's performance by criteria written in their programs, providing reinforcement exercises whenever these seem necessary. But even as machines are now, or are likely to be tomorrow, there is only a poor prospect that they will be able to handle the unexpected difficulties that arise in most people's learning, among which the recurring question "Why bother?" is the most frequent. In one of its rosier passages, the panel does venture the guess that the simulation of the teacher's skills by sufficiently expert systems is not impossible, but for the most part it recognizes that the immediate need is for experiment and investigation to define the bounds of what may be possible.

The belief that the time has come for another wave of interest in new educational technology is even stronger than the panel says. First, and most notoriously, this is a time when few technically advanced societies are able to recruit enough skilled teachers for what are considered essential tasks, teaching mathematics or science more widely, for example. Shortages are especially acute in the field with which the panel is chiefly concerned, information technology. Second, because of the rapidly changing ethos of the high school, students are no longer content to sit and watch teachers make marks with chalk on the walls of the rooms they inhabit but appear (to teachers) to be subversively bent on deciding for themselves what they wish to learn, which is often a recipe for learning (too late, as things are) from their mistakes. Third, there is much more to learn, in circumstances in which the educationalists have not so far been able to devise a curriculum for the modern teenager. Fourth, this is a time when the old advocacy of the cause of continuing education has become an economic necessity; one package of youthful skills no longer lasts a lifetime. How, in these circumstances, can anybody resist this modest committee's appeal for a modest subvention of a programme to find out what computers have to offer students? Not, surely, Mr Kenneth Baker? □

British radioactivity

Sellafield remains at the centre of the storm

LIVING with radiation in Britain has come to require the intellectual stamina for digesting a plethora of technical and often conflicting documents. This is the experience of the past few weeks, which have seen the publication of the first report from the newly appointed Committee on Medical Aspects of Radiation in the Environment (COMARE); the government's response to a critical report on waste management from the House of Commons Select Committee on the Environment; and a further slew of documents from the National Radiological Protection Board (NRPB).

COMARE, which owes its existence to a recommendation of the report two years

ago by Sir Douglas Black about the occurrence of excess leukaemia deaths around the Sellafield reprocessing plant, ironically begins its life by correcting in its *First Report* (HMSO, £4.40) the Black report's calculations.

The unexpectedly high incidence of leukaemia in the neighbourhood of the reprocessing plant, first publicized by a television company, is now accepted as a valid observation which was, in fact, the stimulus for Sir Douglas Black's inquiry. The conclusion of that report was that leukaemia incidence in the neighbourhood of the plant, especially among people born in the 1950s, was indeed greatly in excess of the expected average incidence but not necessarily inconsistent with random fluctuations within the small population concerned.

The committee is however exercised by the discovery that the Black inquiry, which was provided with an official history of radioactive discharges from the Sellafield plant, including that from the fire in the core of a plutonium-producing reactor in 1957, was not told of an incident in which fission products from spent uranium fuel elements had been released to the environment in 1954 and 1955.

The circumstances are odd, and have come to light because of representations by Dr Derek Jakeman, now a reactor physicist at the Winfrith Heath Research Establishment of the UK Atomic Energy Authority (UKAEA), but an employee at the Sellafield (then called Windscale) establishment in the 1950s. Jakeman was one of two employees who called attention to the unsuspected release of radioactivity from the two plutonium reactors on the basis of measurements of soil from his garden near the plant.

Jakeman said this week that irradiated uranium metal fuel elements occasionally incorrectly discharged from the plutonium reactors in 1954 and 1955 would lodge inaccessibly behind a concrete shield and would be allowed to oxidize in the ambient air. He estimated at the time that several kilograms of uranium and associated fission products would have been discharged in oxide form to the local environment. His revised estimate is that a total of 30 kg of uranium and its associated fission products would have been released over a period of perhaps two years; a formal report on Jakeman's reconstruction of this leakage will be published by the Winfrith establishment in the next few days.

What concerns COMARE is that infor-

mation about this release was not provided to the Black inquiry, and that a search of UKAEA archives since carried out did not uncover a reference to the incident. The committee says that the "way in which these data came to light is unsatisfactory and undermines our confidence in the adequacy and the completeness of the available data". The committee nevertheless accepts that the increased radiation doses to young people living nearby in the 1950s are at once less than the natural background and insufficient to account for the leukaemia incidence.

The recalculation of the exposure of the Sellafield population to artificial radioactivity has been carried out by NRPB (*NRPB-R171-Addendum*) in the light of the further information about the uncontrolled oxidation of fuel elements in 1954-55, information about other recorded discharges report to the Black inquiry and new assumptions about the rate of absorption of plutonium and americium in the gut which have recently been recommended by the International Commission on Radiological Protection (ICRP). The effect is to increase the expected incidence of leukaemia as a consequence of discharges from the Sellafield plant by roughly two-thirds, giving an expected number of cases of 0.16 among the study population of 1,255, which is statistically much less than the 4 cases on which controversy continues to centre. NRPB's recalculation also distinguishes between the effects on bone marrow of radiation with high and low linear-energy transfer (LET) in tissues; its report says that although the biological effects of high-LET radiation may have been underestimated, supposing that the degree of underestimation would suffice to explain the four known cases of leukaemia would require that the incidence of leukaemia on account of natural radiation would be much greater than is observed away from nuclear plants.

Sellafield is also the chief focus of the British government's reply to the House of Commons select committee's report, which argued on 12 March that the new reprocessing plant being built at Sellafield to handle oxide fuel from pressurized water reactors should be abandoned. The government now says (*Radioactive Waste*, Cmnd 9852, £3.40) that the financial consequences of abandonment would involve the waste of £600 million already spent on the plant.

More significantly, the government's response also holds robustly to the view that extracting fissile material from spent uranium fuel is a necessary preparation for the next century "when nuclear power will have an essential contribution to make". In other respects, as in its willingness to specify limits for the discharge of gaseous radioactivity to the atmosphere, the government is conciliatory. □

Paying for old lambs

Confusion persists in Britain about arrangements for compensating farmers for the consequences of the ban on the sales of lamb after the accident at the Soviet nuclear power station at Chernobyl. Farmers affected, in North Wales and Cumbria, have been told that the British agriculture ministry will pay compensation to farmers able to show they have suffered loss by having to keep lambs off the market for so long that they no longer qualify for the premium prices paid for lean animals.

But suggestions that the British government will pass the bill, which might amount to £10 million, onto the Soviet government, are premature. That, a spokesman at the agriculture ministry said last week, would be a matter for the Foreign and Commonwealth Office.

Restrictions of the sale of lambs through agricultural markets are being progressively removed, and now apply only to the Lake District region of Cumbria. They were introduced at the beginning of last month because the caesium-134 content of lambs from the affected regions exceeded 1,000 Bq per kg, a limit adopted by the ministry as an "action level" after an *ad hoc* meeting of a European Community advisory committee. The Emergency Reference Level recommended by the National Radiological Protection Board (NRPB) earlier this year is ten times as much, or 10,000 Bq per kg.

As yet, it seems, no British farmer has made a formal application for compensation. There appears to be no plan to compensate farmers for reduced prices from the sales of lamb from farms outside the two affected areas. □

Genetic engineering

Hepatitis vaccine wins approval

Washington

THE US Food and Drug Administration last week approved for the first time the sale of a human vaccine made by recombinant DNA techniques. The vaccine, for hepatitis B, is to be manufactured by Merck, Sharp and Dohme of West Point, Pennsylvania, under the name Recombivax HB, and was developed in collaboration with Chiron Corporation of Emeryville, California.

Recombivax has a potential world market of hundreds of millions of doses for, unlike existing hepatitis-B vaccines, production is not limited by the need for plasma from infected patients. But at \$100 for the course of three injections, the vaccine costs about the same as conventional plasma-derived vaccines and is far too expensive for mass immunization programmes in countries where the disease is most common.

In East Asia and tropical Africa, more than 10 per cent of the population are chronic carriers, and active hepatitis and related cirrhosis are major causes of mortality. Worldwide there are about 300,000 liver cancer deaths per year, most of them thought to result from hepatitis-B infection. In the United States there are 200,000 new cases of hepatitis each year. Those most at risk are offspring of infected mothers, health-care workers, intravenous drug users and male homosexuals.

Although a plasma-derived vaccine, also sold by Merck, has been available in the United States since 1981 only 3–30 per cent of high risk groups have been vaccinated, partly because of fears that the vaccine might transmit acquired immune deficiency syndrome (AIDS). Although Merck says these fears are groundless the company hopes that the recombinant vaccine, which is produced in yeast cells, will prove more acceptable to members of high-risk groups.

Recombivax is made intracellularly in yeast cells that express the hepatitis-B surface antigen. Since early successes were reported in *Nature* (298, 347; 1982), a more complex expression system has been developed, and improvements in expression and scale-up work done by Merck. Patents governing the process are held by the University of California, by Chiron and by the University of Washington. But there is a continuing dispute over some patent rights to the viral genome with Pierre Tiollais' group at Institut Pasteur in Paris, which was the first to obtain a complete viral sequence. The US Patent Office is likely to declare an "interference procedure" to consider the French claims. In the short term, the factor that will limit use of the recombinant product is price. Dr Joseph Melnick of Baylor Col-

lege of Medicine estimates that the price of a treatment would have to come down to around \$1 to be widely used in mass immunization campaigns in the poorest countries, but is optimistic that the potentially unlimited supply of new vaccine will lead to its widespread use.

Dr William McAleer of Merck is also optimistic, believing that the recombinant vaccine made in yeast will in the long term become cheaper than the existing plasma-derived product. McAleer argues that constructing a suitable high-yielding clone has been the most expensive part of development; extraction and purification procedures are inherently cheaper for yeast-derived than for plasma-derived vaccines, because inactivation steps and innocuity testing are less elaborate.

The hepatitis-B antigen particles in

Recombivax are not identical to those from infected humans because they do not include polypeptides coded by the "pre-S" regions of the viral coat gene. In addition, lipid membrane in the particles is from yeast rather than human cells. Nevertheless, about 95 per cent of volunteers immunized in trials developed a strong antibody response and are presumed to be protected. But Chiron has already produced an experimental vaccine that does include pre-S polypeptides and it is hoped that it will provoke immunity in those who do not respond to the Recombivax HB.

Recombivax is Chiron's first therapeutic product to gain marketing approval and the company expects it to earn at least \$100 million a year. The only competitor is sold in Singapore by Smith Kline-RIT, a subsidiary of SmithKline Beckman based in Belgium. But Institut Pasteur in Paris and Genentech in the United States are likely to enter the market and then prices may fall.

Tim Beardsley

UK research grants

Research board pleads for more

EACH year seems to deepen the despair of Britain's Advisory Board for the Research Councils (ABRC). In its annual claim for research funds, now under consideration by the Secretary of State for Education and Science, it finds need to reiterate, in bold type, that scientific research is relevant to "national needs". A series of increases in science spending over the next three years is called for.

The argument the report seems repeatedly to try to drive into the minister's mind is that the new technologies, principally

of scientists from the United Kingdom and in a study from the Science Policy Research Unit that shows the *per capita* expenditure on science in to be the lowest in the industrialized world.

The report protests that everything that can be done to make better use of available money has been done. Dozens of research units have been closed, and more than 3,000 jobs lost to try to give better value for money.

But very little scope remains. If, as the report speculates, it is the government's intention to have a much smaller science base in Britain, there is not even enough money for an orderly transition.

By 1990, the volume of research supported will be some 10 per cent less than at the beginning of the 1980s, despite the value of the science budget having been maintained. The trouble is that the costs of scientific equipment and materials rise much more rapidly than does inflation and this needs to be reflected in the budget.

To end the erosion of research the board asks for an increase over present figures (some £614 million in 1986–87) of £35 million in 1987–88, £50 million in 1988–89 and £60 million in 1989–90. In addition, the Science and Engineering Research Council needs an immediate £9 million to compensate for recent rises in the cost of its subscriptions to international laboratories.

The signs are that the government will be hard to persuade. A white paper (policy document) released last week rejected recommendations from a select committee that the science budget be increased annually by at least three per cent above inflation.

Alun Anderson



biotechnology and information technology, that the government hopes will revitalize the economy, are science-based and crucially dependent on advances in basic research. Demands on the science base, particularly from industry, are thus increasing while the volume of scientific research performed is falling because funds are not available. Evidence of the failure to keep up with other countries in investment can be seen in the brain drain

Human genome

No consensus on sequence

Washington

"CAN it be done? Yes. Will it be done? Yes." With those words, Donald Fredrickson, president of the Howard Hughes Medical Institute, introduced a conference sponsored by his institute on the sequencing of the human genome. But as Fredrickson quickly acknowledged, the questions of how it should be done, when, by whom and who should pay are still up in the air.

What emerged from the conference was a consensus that a physical map of the genome, consisting of an ordered set of cosmids, should be constructed as soon as possible. There is also considerable support for a restriction fragment length polymorphism (RFLP) map.

But few are as enthusiastic when it comes to contemplating the task of sequencing the estimated 3,500 million base pairs in the human genome. "I'm in favor of it", says Cold Spring Harbor director James Watson, "and everyone else at Cold Spring Harbor is against it." The reason is money. Watson thinks there is wide support for starting initial mapping, but many feel that a commitment to determining the sequence will draw resources from other activities.

John Tooze, executive secretary of the European Molecular Biology Organization (EMBO), agrees it would be a disaster if current projects had to compete for funds with the sequencing project. Tooze believes European governments would be hard-pressed to find additional funds, making competition inevitable.

Leroy Hood of California Institute of Technology says it would be a "serious mistake" to jump into a full-scale sequencing project now, as automation will speed sequencing by one or more orders of magnitude in the next few years. Ironically, one of the factors that sparked interest in the sequencing project was the imminent commercial appearance of the automated DNA sequencer that Hood helped to develop (see *Nature* 321, 674; 1986).

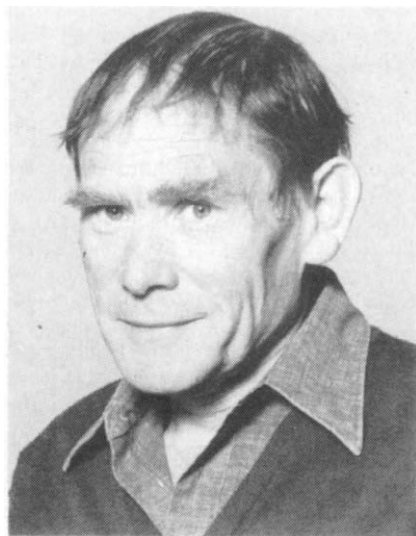
A sequence for the entire human genome, while itself a goal, is only a tool for a better understanding of human genetics, points out Eric Lander of the Whitehead Institute in Cambridge, Massachusetts. Just knowing the sequence tells you almost nothing, argues Lander; it will be useful only when coupled with a better understanding of the structure and function of proteins. Toward that end, Tooze says the European Molecular Biology Laboratory will be putting more of its resources into biocomputing and crystallography.

There has been no dearth of meetings to discuss the sequencing project this year,

and there are more to come. In February, the Howard Hughes Medical Institute conducted a preliminary meeting on the topic in Florida. In March, the Department of Energy hosted a meeting in Santa Fe, New Mexico, and in June the topic was the focus of a Cold Spring Harbor symposium (see *Nature* 322, 11; 1986). Next month, the National Academy of Sciences will air its views at a meeting in Woods Hole, Massachusetts, and in October the National Institutes of Health will discuss what their role in the project ought to be.

New career for Sydney Brenner

Dr Sydney Brenner, director of the Medical Research Council (MRC) Laboratory of Molecular Biology in Cambridge, is soon to have a chance to give up administration and pursue his own personal research. This week, MRC announced that it has set up a new Molecular Genetics Unit which Brenner will head from 1 October this year until his retirement. The unit will be a small one, of 4 or 5 researchers, housed within the MRC Centre in Hills



Sydney Brenner

Road, Cambridge. Four hundred thousand pounds will be provided to establish the unit and a budget of £200,000 a year awarded thereafter. Research at the unit will concentrate on areas in which Brenner has already won international renown: methods of gene mapping; the characterization of genes affecting the development of the nervous system in *Caenorhabditis elegans*; and aspects of gene evolution. Dr Aaron Klug, who won the Nobel prize for Chemistry in 1982, will succeed Dr Brenner as director of the Laboratory of Molecular Biology.

Alun Anderson

Enthusiasm for starting the sequencing project continues to run high at the Department of Energy. The department has already started constructing a physical map of the human genome, as well as supporting research on database management and automated data entry. But its cost estimates for the sequencing project have dropped considerably during the past year — down by an order of magnitude from the \$3,000 million that was thought necessary at the March meeting.

In spite of uncertainty about how best to proceed with the project, Sydney Brenner, director of the Medical Research Council Laboratory of Molecular Biology in Cambridge, United Kingdom says that everyone interested in medical research ought to be interested in the sequence of the human genome. Those that are not, says Brenner, "have to be considered a subject for appropriate treatment".

Joseph Palca

David Swinbanks in Tokyo adds: It would take US \$600 million and 30 years to sequence the entire human genome by a "sequencing factory" according to Akiyoshi Wada, professor of biophysics at Tokyo University. While that may seem a lot, he estimates that if it was done by 100 experts it would take \$1,100 million — and 250 years.

Since 1981, Wada's group has been trying to develop an automated DNA sequencing system in collaboration with industry under the support of the Science and Technology Agency. A robot that performs DNA fragment separation according to the Maxam-Gilbert or Sanger M13 protocol (see *Nature* 307, 193; 1984) has been developed with Seiko Instruments and Electronics Co. And Toyo Soda Manufacturing Co. has designed a specific DNA extractor with high base resolution, while Fuji Photo Film Co. has created a system for mass production of films for gel electrophoresis.

Now Seiko and Hitachi Software Engineering Co. are competing in development of machines to read the gels and type out the sequence. By combining these operations, Wada estimates that it should be possible to sequence one million base pairs per day.

Once a working system is developed, Wada thinks that sequencing should be extended to all parts of the genome, not just those of immediate interest. He compares the operation with geographic mapping, which is carried out irrespective of the "profitability" of the area to be mapped.

If a sequencing factory can be built, Wada emphasizes that it would not be "Japan Incorporated" against the rest of the world. He wants an international centre that would be open to scientists of all nationalities and intended for the benefit of all mankind. □

Biotechnology

Regulations please nobody

Washington

CONTROVERSY continues in the United States over the federal government's proposals to regulate biotechnology. At a recent congressional hearing, representatives of the American Society for Microbiology, the Environmental Law Institute and the Ecological Society of America all criticized the administration's plan to exempt from high-level review environmental releases of intergeneric organisms whose introduced genetic material is "well-characterized and contains only non-coding regulatory regions".

The administration's latest proposals on biotechnology regulation were published after much delay on 26 June. They recognize that non-living products of genetic engineering technology should be regulated in the same way as conventionally made products, but require that environmental releases of engineered organisms containing DNA from different genera or from a known pathogen should receive high-level review. The exemption for transferred non-coding regulatory sequences applies whether or not the donor organism is a known pathogen.

The proposals were examined last week at a joint hearing by three subcommittees of the House of Representatives' Science and Technology Committee. Monica Riley, chairman of the American Society for Microbiology, said that "strain construction that places a strong, effective regulatory sequence in a position of control over any particular gene can increase the expression of that gene many fold", thus affecting a microorganism's capacity to compete. Riley also pointed out that regulatory sequences carry specific determinants that turn genes on or off; changing the signals to which a gene responds could constitute a substantial change in the biology of the recombinant organism.

Some congressional staff speculate that this criticism from respected professional bodies might persuade the administration, which is anxious not to overburden the emerging biotechnology industry, to remove the exemption. Industry organizations publicly support the administration's proposals, but some privately concede that the non-coding exemption could undermine public support and they would rather have it removed.

Meanwhile, the administration's plan has come under attack from a more predictable quarter. Jeremy Rifkin of the Foundation on Economic Trends has sued the government for omitting to conduct an environmental impact statement and failing to keep an adequate record of its development.

Tim Beardsley

West German nuclear power

The fast breeder is stumbling

Hamburg

REIMUT Jochimsen, Social Democratic Economy Minister of Nordrhein-Westfalen, has put aside political questions and called for a halt to the fast-breeder programme on scientific grounds. Fulfilling his earlier promise to keep scientific and political issues separate, he has published an expert opinion (roughly 100 pages long) written by co-workers in his ministry, where grave doubts are raised about the safety of the 300 MW prototype reactor. The problems are enough, Jochimsen emphasizes, to make the fast breeder the most expensive misinvestment ever in industrial history.

The report clearly shows that politicians have been careless in trusting the calculations of engineers from the reactor building and operating companies, in particular those of the Schnellbrüterkraftwerksgesellschaft (SBK), on which many of the 17 permits already granted were based. Under certain very adverse conditions, a runaway reaction could take place. In contradiction to earlier claims, there are serious doubts about the breeder's safety.

In 1984, experts of the companies involved said in court that a disastrous failure of the ventilation system was impossible. This expert opinion proved wrong when after less than two years of routine test running (without nuclear fuel cells) two of the ventilation devices failed. The experts also excluded absolutely the possibility that oil could enter the radioactive primary system of the reactor. But that happened. Never, they said, could water and sodium, the reactors' primary coolant, meet. But that too has happened.

Parts of the reactor, which must never fail, are called "holy" by the technicians. But some of them have been found to be not safe under any conditions. A grating to hold the fuel cells, for example, can rust. Jochimsen said that for such parts a "repeatable control" is necessary, which is impossible in the planned prototype. Rust has also (as long ago as 1980) been found in the reactor tank, and technicians have found weaknesses in weld seams. The method of estimating the probability of a runaway accident was not scientific, the report says. All the Gesellschaft für Reaktorsicherheit (the company responsible for reactor safety) did was to ask selected experts if they thought such an accident likely. Another analysis, carried out at the Kernforschungszentrum Karlsruhe, was found to be incorrect due to a faulty computer program.

Opponents of the Jochimsen report, such as the leader of the Nordrhein-Westfalen Christian Democrats, Kurt Biedenkopf, say that its arguments are

false. The truth, they say, is that Jochimsen is executing the Social Democratic Party's decision to stop the nuclear programme. A representative of the Ministry of Research and Technology said that Jochimsen's political intentions are evident because the minister published his paper only 11 days after the Nordrhein-Westfalen members of the Social Democrats had decided to stop the breeder. This accusation is hard to believe, however, for the report cannot have been produced so quickly.

Critics have not attacked the technical arguments presented. Jochimsen points out, but only produced political arguments. The head of SBK, August Wilhelm Eitz, argues that all necessary modifications will be made. But that still leaves one key objection from Jochimsen: in the foreseeable future, there is no acceptable place and method for disposing of the nuclear waste from the reactor.

The Social Democrats now want to go further. In Nordrhein-Westfalen, permission will not be given for new sites for nuclear plants. The party is working on a bill to be introduced into the Bundestag in the autumn. According to representatives, it demands that no nuclear plants at all should be permitted to run in West Germany, even those already built. Preferential treatment for nuclear energy should be ended. In addition, the Social Democrats will introduce clear standards of radiation analysis and protection to avoid the confusion experienced after the Chernobyl disaster. The law, which could come into force if the Social Democrats win next year's federal elections, would also automatically terminate the DM7,000 million breeder. The Nordrhein-Westfalen government cannot enforce a decision on its own because the federal government can make it grant permission for the fast breeder reactor.

Heinz Riesenhuber (CDU), Minister of Research and Technology, has already mentioned that option in an interview, but he knows that it is unlikely to be successful. On the other hand, if he abandoned the project he could save more than DM200 million. The problem is very delicate and the legal situation unclear. If the project is stopped without definite technical reasons, the power supplier will get back the DM1,400 million it paid. Belgium and the Netherlands will also want their DM1,600 million returned. If, however, Jochimsen's paper is telling the truth, the reimbursement need not exceed the present value of the plant, which would not be very high, given a plant too dangerous to run. A meeting planned for September should make things clear.

Jürgen Neffe

Atomic power

No nuclear doubts for Japan

Tokyo

By 2030, Japan will have more than 120 nuclear power reactors supplying over half the country's electricity needs. Fast-breeder reactors will be on line, and fuel cycle facilities will provide a large proportion of the nuclear industry's enriched uranium requirements, reprocess most of its spent fuel and store all its waste. Such is the "vision" of Japan's nuclear industry outlined in a report recently submitted to the Ministry of International Trade and Industry by the nuclear subcommittee of the Advisory Committee of Energy. But where will all these nuclear plants be located and will the Japanese public willingly accept them?

Japan has 32 reactors generating nearly 25 million kilowatts, putting Japan behind the United States, the Soviet Union and France in total nuclear capacity. Twenty-six per cent of Japan's electricity is nuclear generated.

According to the report, nuclear power capacity will more than double to 62 million kilowatts by the end of this century and, assuming a growth in Gross National Product of 2.5 per cent, rise to 137 million kilowatts by 2030 to provide 58 per cent of Japan's needs. This will require the construction of 116 new reactors at the rate of 2 or 3 a year. Ten new reactors are under construction and seven more are planned.

Two areas in Fukui and Fukushima prefectures contain most of the nation's nuclear plants. Their presence has brought economic benefits but there is resistance to further expansion. The town of Ohi in Fukui Prefecture, for example, has earned more than ¥11,000 (£47 million) in fixed property revenue funds, grants from the central and local government and redemption taxes during the past 15 years. But plans to double the number of reactors in the town to four have led to mass demonstrations; in 1984 1,400 riot police had to be mobilized to contain thousands of demonstrators protesting outside public hearings in the town.

The subcommittee's report makes several recommendations to promote public understanding of nuclear power so that expansion can proceed "smoothly". Safety should be strictly maintained in the plants to ensure the public's understanding of the "safety" of nuclear power. Nuclear power technology should be explained in simple, easy-to-understand terms, and the central and local government should positively promote nuclear power in school education.

Japan certainly has an enviable safety record, with far fewer shutdowns due to accidents than for example, the United States. But accidents do occur, such as the leak of low-level radioactive waste at the

Tsuruga power plant in 1981 and the exposure of an official of the International Atomic Energy Agency to traces of plutonium earlier this year.

A key element in future plans is the construction of a huge fuel cycle facility for uranium enrichment, fuel reprocessing and low-level waste storage in Aomori Prefecture at a cost of one million million yen (£4,300 million). But this facility, for which approval was granted by the prefectural government in 1985, has yet to gain the unanimous approval of the local fishing cooperative and the start of construction has been delayed.

Fishing cooperatives wield considerable political power in Japan as do other blocks of rural voters who traditionally support the ruling party. Promises of ¥177,000 million yen (£750 million) in contracts for local companies and 1,400 jobs have won approval for the project from other sections of the local community.

If the Aomori complex is completed on schedule in the 1990s, it will provide 30 per cent of Japan's enriched uranium requirements by the year 2000, and will be capable of reprocessing most of the spent

fuel that is now sent to French and British facilities. According to the report, a second fuel reprocessing plant will be built in Japan in 2010 to meet increased demand.

Low-level radioactive waste will be stored at Aomori in a plant with an eventual capacity of 3 million 200-litre drums. It will be capable of dealing with all Japan's low-level waste produced by 2030 and beyond.

Greatest growth over the next 45 years is seen in establishing the fuel cycle. Expenditure is expected to amount to ¥70,000 million million yen (£300,000 million) by 2030. The only note of restraint in this bullish assessment of the future of Japan's nuclear industry is in regard to fast-breeder reactors, commercial operations of which are predicted to start in 2020, ten years later than expected.

But a big question mark hangs over the locations of the proposed new plants. The report calls for the establishment of 20 new sites without clearly stating where they will be. Development of new siting technology is advocated so that nuclear plants can be placed underground, on the sea, and close to population centres. Clearly Chernobyl has not shaken the confidence of Japan's nuclear power planners.

David Swinbanks

Chernobyl

Experimental speculations

THE Soviet Politburo, apparently responding to a torrent of questions about the mysterious "unauthorized experiment" said to have been under way when the Chernobyl reactor accident occurred on 26 April, said last week that the power station crew was attempting to operate the reactor at a low power, below that for which it was designed.

From the beginning, Soviet statements have said that the Chernobyl reactor was operating at 7 per cent of full power at the time of the accident. Many reactor engineers in the West now consider that the most likely explanation of the Soviet statement is that those concerned were trying out some scheme for operating the reactor's housekeeping functions from the station turbine proper, and not from the stand-by diesel sets provided to guard against disconnection from the Soviet electricity grid.

While operating at full power, reactors are naturally able to deal with this "hotel" load from their own production of electricity, but the removal of fission product heat from a shutdown reactor requires an external source of power. Measurements of fallout from Chernobyl suggest that the reactor had been shut down for some hours before the first release of radioactivity began, which would suggest that the decay heat of the fission products

would by then amount to some 4 per cent of the total power, less than the 7 per cent quoted by Soviet sources.

Dr David Hicks, director of the UK Atomic Energy Authority's water reactor programme (chiefly concerned with the safety of pressurized water reactors) said earlier this week that such an operation, well outside the power range for which the Soviet RBMK reactors are designed, could well have made the Soviet reactor prone to the instabilities for which it is known, especially because a core which had been in use for 1.5 years would have accumulated such a load of fission products that the production of even a few per cent of full power by fission would have required that the control rods should be to a large extent withdrawn.

The occurrence of steam voids within the cooling system would, in these conditions, have created power excursions which it might have been difficult to control by means of the control rods, which move in narrow water channels where their free fall is impeded. Hicks emphasized, however, that his theory of what may have happened is only one of many.

Much of what is known of the Soviet reactors derives from three consultations in the mid-1970s when British and Soviet designers exchanged information about the RBMK and its British analogue. □

French budget

Bleak prospect for researchers

THE French government research budget for 1987 is now more or less complete. No figures have been released, nor are they likely to be before September, but the indications are that the spending power of French scientists will at best be pegged at around the levels of 1985. There may also be redundancies, as the Prime Minister, Jacques Chirac, has ordered a 1.5 per cent across-the-board cut in government employment. There is hope, however, that the job losses will be partly compensated by a growth in the recruitment of young scientists, under a government scheme to counter youth unemployment.

The worsening prospects for French scientists, coming after the boom years of the early 1980s, were heralded by an 8 per cent cut in the science budget made in April, just a month after the new government came to power. The change shows

how depressingly shallow-rooted was the science policy-making system instituted by the previous administration during its five years in power. Research, it seems, never became a proper political and administrative fiefdom, and now the parts of the old estate, the ministry of research and technology, have been scattered among the ministries from which they were first confiscated in 1981.

The result, according to one senior French scientist, is that there can no longer be a science policy in France. Rather, research will get the scraps of the budget that are left after the real political carnivores — the ministers responsible for police and defence, among others, in this government — have torn at it. Only then will the minister of research and higher education, the well-meaning physicist Alain Devaquet, have a chance to form a policy. The importance for research and technology of having a politically powerful champion (in the previous French government it was Jean-Pierre Chevènement, who led the largest single wing of the socialist party) has never been clearer. Without power, the budget determines the policy; with it, the reverse is possible.

The Conseil Supérieur de la Recherche et de la Technologie (CSRT), the independent group of high-level scientific advisers to Devaquet, is already well aware of the sea-change facing French science, but seems powerless to act. Having seen the outline budget for 1987, the Conseil has been forbidden by Devaquet to speak of it. The minister will present his budget, which given inflation and the April cuts would need a 6 per cent increase in current francs to retain the spending power of 1985, probably in September. Only then will CSRT, of which Devaquet is *ex-officio* chairman, be free to publish its "advice".

Nevertheless, François Kourilsky, the Marseilles immunologist and scientific president of CSRT, has dared to warn that any suppression of scientific jobs would have "long-term effects incommensurate with the economies achieved", and to demand that the government "finish and publish a research and technology policy". This may come with Devaquet's awaited presentation, but the feeling is now clear in France that science has returned to the doldrums of the 1970s. Then, ministers in the Délégation Générale à la Recherche Scientifique et Technique (DGRST) who were responsible for "coordinating" the research budget in the prime minister's office were renowned for their fine forward-looking speeches in the Assemblée Nationale — but for their almost total lack of power to influence real events.

Robert Walgate

US-Japan trade

No easy end to chip war

Washington

AN agreement is expected early this week in the long and increasingly bitter negotiations between Japan and the United States on semiconductor trade. The United States is seeking greater access to Japanese markets for US semiconductor manufacturers, and an end to "dumping" of Japanese semiconductors on international markets. As both legal and self-imposed deadlines loom, negotiations have entered what a spokesman for the US semiconductor industry termed "a period of trench warfare".

For more than a year, the US semiconductor industry has been pursuing a trade complaint against Japan over access to Japanese semiconductor markets. US chips now account for only 10 per cent of sales in Japan according to US industry figures, as opposed to 83 per cent of sales in the United States and 55 per cent of sales in Europe. In late May, US Trade Representative Clayton Yeutter worked out a framework for concluding the trade case, known as the 301 case after the section of the Trade Act of 1974 that covers unfair trade practices. Included in the framework is the resolution of two other complaints that Japanese companies are selling erasable programmable read-only memory chips (EPROMs) and 256 kilobyte dynamic random access memory chips (256K DRAMs) below their production costs. A preliminary judgement made last year by the Department of Commerce found Japanese companies guilty of dumping in violation of international trade agreements. But in early July, both sides reached a tentative agreement to drop the dumping cases so that negotiations in a comprehensive agreement could continue. Since June, US and Japanese negotiators have been trying to forge an agreement under Yeutter's framework, but so far without success. Without a final decision to stop the dumping cases this week, permanent dumping duties will be imposed on all imported Japanese EPROMs and 256K DRAMs.

Opening Japanese markets to US goods is not straightforward. Despite the power of the Japanese Ministry of International Trade and Industry, it cannot force Japanese companies to buy US chips. The Japanese government has offered to open a special office to assist foreign companies in making semiconductor sales.

Japanese semiconductor producers are large, vertically integrated companies that can subsidize losses in their semiconductor division to purchase market share, an option not open to generally smaller US firms. No agreement will

Terrorists strike again

Hamburg

THE terror against technology in West Germany continues. Following the killing of Siemens research manager Karl Heinz Beckurts by terrorists of the Rote Armee Fraktion (RAF), a bomb exploded at the Fraunhofer Institute for Laser Research in Aachen at 5 a.m. on Thursday last week. The total damage has not yet been estimated, but some sensitive machines seem to have been affected. In a letter found near the site, a "Fighting force Sheban Atlouf" admitted its responsibility for the attack. The institute, which is part of the Technical University of Aachen, is carrying out research only on the introduction and use of laser techniques for industrial and medical purposes, according to the head of the institute, Professor Herziger.

RAF struck again last Friday at 5 a.m. This time their chosen target was more in line with their announced aim of hitting military and nuclear research institutes. The bomb exploded outside the headquarters of Dornier GmbH, owned by Daimler Benz, in Immenstaad. Only a wall and about 250 windows were damaged.

Police think that this may be the beginning of a series of bombings against what RAF calls the "Militärisch Industrieller Komplex". Last year, when some bombs exploded near institutes for biotechnology and gene research, the new aim of the terrorists became clear. In the underground newspaper *Sabot*, anonymous sources have announced that "harmless" people in various institutes will face assassination. Protection for such a wide range of imperilled people and institutes cannot be given. Scientists have a hard challenge to face.

Jurgen Neffe

change this situation.

A document leaked to the *Financial Times* of London gave details of a working draft of the agreement. The United States would agree to drop its dumping cases, and Japan would try to increase US market share in semiconductors to 20 per cent. To prevent future dumping disagreements, the United States would request

WITH a self-imposed deadline of Saturday, 26 July for concluding the negotiations, Japanese representatives were no doubt surprised when the US team announced on Friday that negotiations had to stop at 1 p.m. The reason? The Trade Representative's office had scheduled its annual picnic for that day, and not even Washington's current heat wave was going to stop plans for a softball game during the picnic.

immediate consultations between the two governments. These consultations would be limited to 14 days, in contrast to the year it now takes to resolve dumping issues. If dumping is suspected, the Japanese government will encourage its industry to provide documentation for the legitimacy of its sales figures.

Even if an agreement is reached, chip prices in Japan will probably stay comparatively low, and US importers are likely to find "loopholes" — purchasing semi-finished goods containing chips or buying chips at Tokyo discount stores. Nevertheless, Japanese companies do appear to be worried by the agreement. They find it blatantly unfair and one-sided, placing Japan at a disadvantage to South Korea, Taiwan and especially to US companies that will have access to Japanese production costs under anti-dumping agreements. Japanese industry sees this agreement as being largely beyond their control, having arisen out of the politically close relationship between President Reagan and Prime Minister Nakasone.

As talks moved into the eleventh hour, both sides seemed intransigent. Three US manufacturers complained formally to the Department of Commerce that Japanese companies were trying to cut last minute deals at bargain prices with US customers before signing an agreement. By 30 July a decision has to be made on the EPROM dumping case, and the next day President Reagan must announce what action he will take in the 301 trade case. As *Nature* goes to press, it is impossible to say what the likely outcome would be.

The US industry is frustrated by the current negotiations. It believes it has an open and shut case against the Japanese both for dumping and for closing their markets. Hopes for agreement may have given way to a desire for retaliation against Japan for its trade practices. But a comprehensive agreement is still the ultimate goal. At stake could be nothing short of the future of the US semiconductor industry. **Joseph Palca & David Swinbanks**

High-technology trade

Anger over supercomputer veto

OFFICIALS at the British Embassy in Washington hope that they can soon resolve "without major compromise" the difficulties of British academics in obtaining US-manufactured supercomputers. Intensive discussions have been taking place over an issue that last week sparked a debate in the House of Commons.

The trouble began when the University of London Computer Centre found that its purchase of a Cray I supercomputer had been blocked by US government regulations. A condition of sale of the supercomputer is that scientists from Eastern Bloc countries and China must not gain access to the machine. But the centre's director, Richard Field, says that although the scientific standards of work carried out on the machine will be monitored, it is not possible to police all the computer's users to ensure there are no scientists from proscribed countries among them.

What has really enraged British opposition politicians is that the computer the centre wants is second-hand and already in Britain; all that has been requested is its transfer from the Atomic Energy Authority's Harwell laboratory. That the United States can control the movement of a computer inside Britain is described by Liberal Member of Parliament Mr Paddy Ashdown, who initiated the House of Commons debate, as a "flagrant breach of British sovereignty". And it became clear during the debate that while this may be the first time US controls have hit academics, British high-technology companies have had difficulties with US government regulations for some time. Indeed, Ashdown claims that one British company was effectively driven out of business by US sanctions.

The US regulations come on top of the COCOM agreement that regulates the export of strategically sensitive goods from members of the North Atlantic Treaty Organisation and Japan. British companies that wish to export manufactured goods containing any US component (as do most electronic products) now find themselves having to apply for a licence from both the UK Department of Industry and the US Department of Commerce.

Ashdown's exhortations to take on the United States did not go down well with the government. The official response, from Minister of Information Technology Mr Geoffrey Pattie, was that "if we could compel the United States to withdraw . . . the problem could be eliminated easily, but we cannot . . . and US reaction to a direct and comprehensive challenge . . . cannot be predicted". Rather than accept the opposition view of British Prime Mini-

ster Margaret Thatcher's "craven subservience to President Reagan". Pattie dwelt on the overall advantages to Britain of commercial and scientific links with the United States — plus the new efforts to reduce that dependence by strengthening links with Europe in large-scale development programmes. Problems with the United States have to be dealt with by a "case-by-case" approach, he says.

This view is scarcely welcome to British Embassy officials in Washington who have to do the case-by-case negotiations and hope for a more general understanding. An agreement is needed soon if research at the University of London is not to be severely disrupted. The computer should be delivered next month according to the contract, on which a prepayment has already been made. If the computer is not in place by the end of the year, the contract will be void and the computer centre will have to do what it can to seek redress. Any decision will also have a bearing on a contract for delivery of a Cray to the Science and Engineering Research Council's Rutherford Appleton Laboratory next year. **Alun Anderson**

Science Digest dies

Washington

SCIENCE Digest last week became the latest casualty of the turmoil hitting popular science magazines. Hearst Magazines, its publisher, announced that the September issue of *Science Digest* will be the last. The subscription list and licensing rights to the name have been bought by Time Inc., publisher of *Discover*, for an undisclosed sum.

The announcement came as a shock to editorial staff, who have been given a week to clear out their desks. Hearst had been trying unsuccessfully to sell the magazine for several months before finally deciding to cease publication, mainly because of a drastic fall in advertising revenues (see *Nature* 322, 99; 1986). Hearst expects to retain and reassign about a third of the 26 *Science Digest* staff.

Time Inc. also recently bought for \$6 million the logo and subscription list of *Science 86*, the popular science magazine published until last month by the American Association for the Advancement of Science; its readers will be offered *Discover* instead. *Science Digest* subscribers will be offered a choice of various Time publications. A spokesman for Time Inc. said that Hearst had already made the decision to cease publication of *Science Digest* before Time agreed to buy the subscription list: the *Science Digest* logo is not likely to be used.

Tim Beardsley

CORRESPONDENCE

GAP protest over South Africa

SIR—We wish to bring to the attention of the scientific community the circumstances that led to the cancellation of the Third International Workshop of the Group for Aquatic Primary Productivity (GAP) scheduled for 29 April–4 May in Durban, South Africa.

GAP was established in 1980 under the aegis of the International Society for Limnology and the International Association for Ecology to provide a forum for oceanographers and limnologists studying theoretical and methodological aspects of primary production in marine and fresh waters. Successful GAP Workshops were held in Konstanz, West Germany, in 1982 and Haifa, Israel, in 1984.

When the International Committee of GAP accepted the invitation of our South African colleagues to hold the third workshop in that country, we were aware of potential problems. It was clear that some scientists, as individuals, would choose not to attend an event in a country whose governmental racial policies are morally repugnant to most, if not all, of us. This viewpoint is understandable and such a decision is the prerogative of each individual.

We did not expect, however, to be faced with a situation in which a significant number of our colleagues who had expressed their intention to attend the workshop were expressly forbidden to do so by their superiors or administrations. In at least one case, a young scientist was threatened with jeopardizing his future career if he participated in any capacity at the GAP meeting. As a result of such pressures, the number of prospective participants in the workshop became so limited that we were reluctantly forced to cancel the meeting.

GAP is a scientific organization affiliated to the International Council of Scientific Unions (ICSU). For many years, ICSU has striven to guarantee the freedom of *bona fide* scientific exchange. The Principles of Universality as formulated by the ICSU Standing Committee on the Free Circulation of Scientists are basic to all ICSU objectives and have been reaffirmed by the latest ICSU General Conference in Ottawa in October 1984. Actions that deny scientists' freedom to travel to legitimate scientific activities are a flagrant and dangerous violation of these principles.

Science is an international endeavour, unfortunately tainted by political realities. It behoves the scientific community to be vigilant in maintaining its rights of free exchange and circulation. The forced cancellation of the GAP Workshop has no impact on the South African government's racial policies and only serves to increase the isolation of our colleagues

who are largely counted among the liberalizing forces there. We can only caution the scientific community that, alas, this will not be the last encroachment upon the Principles of Universality and advise vigorous and early protest in future instances.

TOM BERMAN
(Chairman, GAP International
Committee)

*Israel Oceanographic
& Limnological Research Ltd,
The Yigal Allon Kinneret
Limnological Laboratory,
POB 345,
Tiberias, Israel*

RGO move OK

SIR—The decision of the Science and Engineering Research Council (SERC) to move the Royal Greenwich Observatory (RGO) to the University of Cambridge is a historic decision, whose consequences are not fully realized by *Nature*. Your leading article (*Nature* 322, 1; 1986) belittles the SERC choice as being based on the unwillingness of RGO staff to travel north of the Trent. This point-scoring ignores the facts that ex-RGO staff hold positions at the Royal Observatory, Edinburgh (ROE) — there has been a good history of staff interchange between RGO and ROE — and that large numbers of the RGO staff spend a large proportion of their time working and living overseas on La Palma and elsewhere. Jerry Sellwood's pro-Manchester letter (322, 106; 1986) comments on the decision by raising issues irrelevant to astronomy, such as national policies of regional development and an alleged prejudice by everybody against Manchester.

In deciding to move RGO to Cambridge, SERC is consolidating one of the world's great centres for astronomy, bringing radio, theoretical and optical astronomy together with the instrument science and engineering on which the progress of astronomy depends. It is true that many RGO staff (including ourselves) and many British astronomers opposed the move, and that the University of Sussex's astronomy programme could be damaged, an issue that SERC should address in the future. But we believe that of the choices SERC set itself the decision to move to Cambridge is best for British astronomy as a whole.

We hope that the Secretary of State for Education and Science, taking proper account of the financial arguments, will stick with the clear scientific decision of SERC. Further vacillation would prolong the problems for RGO and British astronomy; surely astronomers in the United Kingdom now need to put this matter be-

hind them and move confidently into the twenty first century, getting on with the tasks at hand.

BOB ARGYLE, CHARLES JENKINS, DEREK JONES, ROBERT LAING, TOM MARSH, BILL MARTIN, PAUL MURDIN, MAX PETTINI, HANS SCHILD, KEITH TAYLOR, ROBERTO TERLEVICH, KEITH TRITTON, JASPER WALL
*Royal Greenwich Observatory,
Herstmonceux Castle,
Hailsham, East Sussex BN27 1RP*

Japanese education

SIR—I would like to add to Alun Anderson's News item (*Nature* 321, 6; 1986) on Japanese education from the viewpoint of an undergraduate student. The University Council recently decided to increase the opportunity for applicants to take the university entrance examination with the aim of increasing the flexibility of the universities. But I wonder if it will not produce another worse result; to standardize the quality of students and to rank more clearly all of the universities. The educational industry will also gain power, worsening the situation we call "the examination war ordeal". Even now, many applicants choose their universities only for their name value, and after entering have no clear object of study. To solve these problems, the universities must be given more independence and autonomous rights.

The PhD glut ("overdoctors" or unemployed PhD holders) is another difficult problem. If one wants to get a post in a university, one must put up with years as an overdoctor because of the low staff turnover. Otherwise one must move to private enterprise. For this reason, I think, Japanese basic research cannot develop fully.

SHINJI HIRANO

*Faculty of Science,
Kyoto University,
Kitashirakawa, Sakyo-ku,
Kyoto 606, Japan*

Alun Anderson replies: I agree with Mr Hirano. All state universities have traditionally held their examinations on the same day, effectively preventing anyone from applying to more than one state university in any one year, a situation that commits candidates to the uncertain task of selecting a university with a pass mark thought likely to match the candidate's likely examination mark. Universities are now to be divided into groups with examinations on different days. But this may not solve the problem of universities forming a hierarchy according to their pass marks. Kyoto University law school, for example, fears that in future all its candidates will take both Tokyo and Kyoto University law school examinations — and all those accepted for Tokyo will go there. That could have the immediate effect of converting Kyoto into a "second-rank" institution. □

Catching atoms in beams of light

The past few weeks' excitement about the trapping of atoms in laser beams is well justified; the applications of the technique could be exciting.

THE notion that it might be possible to trap an atom in a single beam of light is, of course, nonsensical. Atoms capable of interacting with light of specified frequency will sense what is called radiation pressure, and be propelled along in the direction of the beam. This is how newly formed stars clear a nearly empty space in their immediate vicinity and, for that matter, how the outer regions of all stars are prevented from gravitational collapse upon themselves. It is only natural that people should not hitherto have sought to trap single atoms in a single laser beam.

That diffidence will now have to be dispensed with. A group from AT&T Bell Laboratories, in a continuation of work by A. Ashkin going back to the early 1970s, has indeed been able to trap a large collection of sodium atoms in a single laser beam. The group now reports (*Phys. Rev. Lett.* **57**, 317; 1986) that it has been possible to gather as many as 500 atoms into a volume no more than 10 micro-metres in dimension, which corresponds to the respectable density of 10^{11} atoms per cm^3 .

But how can such a feat be accomplished? Steven Chu, J.E. Bjorkholm, A. Ashkin and A. Cable have been able to start from a great deal of expertise accumulated over the years at their own establishment as well as from the many successes in recent years with different kinds of traps for atoms, usually built around combinations of electrostatic and magnetic forces. Devices of that kind, mostly elaborations of the Penning trap, have been widely used in recent years for the spectroscopy of isolated, or nearly-isolated, atoms.

The essential requirement for success appears, however, to have been the suspension of conventional belief. In a companion paper (*ibid.* **57**, 310; 1986), J.E. Pritchard *et al.* rehearse the case for believing that optical trapping is impossible — before showing that not to be the case. The principle is simple enough. In a coherent beam of light, the flux of momentum is represented by the Poynting vector, which is proportional to the vector product of the electric and magnetic field intensities (and thus directed in the direction of an ordinary light beam). By an extension of the nineteenth-century Earnshaw's theorem in electrostatics, it then turns out to be the case that the Poynting vector must be divergenceless across any closed surface that does not contain a source or sink for radiation, just

as the divergence of the electrostatic force must be divergenceless over a surface containing no electrostatic charges.

In the optical case, what this seems to imply is that the force on irradiated atoms, which must be in the same direction as the Poynting vector, will also be divergenceless and thus could not be formed in such a way as to be directed inwards everywhere on some closed surface. Ironically, say Pritchard *et al.*, Ashkin is one of those to have argued that Earnshaw's theorem is a bar to laser trapping for atoms, thereby "discrediting these proposals and discouraging any others". One way round the difficulty would be to use light intensities varying with time, a technique that works with radio-frequencies; a temporarily inward array of forces can, for example, be switched off when a sufficient number of atoms have been corralled into some other kind of trap. Doing this at the frequency of optical radiation would be a different matter.

The essence of what Ashkin's group has now done is to use a strongly focused laser beam whose intensity profile is far from uniform but, rather, gaussian, with the peak intensity along the centre. Moreover, the laser generating the beam is tuned to a frequency some way below the natural resonance frequency of the sodium line, with the result that the force on sodium atoms can be controlled at an arbitrarily low level. In these circumstances, the only forces on sodium atoms that matter are those between the electric field of the laser beam and the electric dipole moments which they induce in atoms exposed to them. The forces depend on the detuning of the laser beam, but not as critically as the scattering.

The result is that sodium atoms can be made to move towards the most intense part of a non-uniform laser beam. To be sure, the magnitude of this effect is greater as the frequency of the laser approaches the resonance frequency (which is the condition when scattering could be a nuisance), which implies that a trade-off must be struck between the closeness of the approach to resonance and the steepness of the gaussian profile. That, says Ashkin's group, accounts for trapping on the axis of the laser beam. Trapping in the axial direction depends on the rapidity with which the beam is focused in space.

The actual success reported by Ashkin's group hangs crucially, as is acknowledged, on the way in which it has been possible to

start with sodium atoms "cooled", again by the use of laser beams, to an effective temperature of a quarter of a millikelvin. The idea there, implemented in the past few years, is that it is possible to rob an atom of its translation velocity in the direction of a beam of light whose frequency is well-defined by means of the transfer of momentum that occurs on scattering. The trick is to arrange the difference of the laser and the resonance absorption frequency of the atom so that the Doppler shift works in the direction of cooling, but does not as easily allow the transfer of momentum to atoms which are already nearly at rest. But if this works in one dimension, why not arrange that three lasers at right angles should bring a whole package of atoms to near-rest? This arrangement, called "optical molasses" for obvious reasons, was demonstrated only last year. In the new experiments, a package of atoms trapped in such molasses is used as the sources of material to be trapped by the single convergent laser beam, suitably detuned.

For the time being, everybody seems too breathless to wish to speculate where the new development will lead. It is nevertheless remarkable that the lifetime of a package of some hundreds of sodium atoms in such a trap can be as much as several seconds — virtually aeons by the frequency of the resonance absorption line. The lifetime of the trap is limited, for practical purposes, by the residual scattering of light by the atoms in the trap, from which they gain energy and an increased capacity to escape. One little refinement is that, according to Ashkin's group, it is possible to move the package of trapped atoms bodily from place to place, which is why AT&T Bell Laboratories, public relations people have coined the term "optical tweezers".

So what happens next? The insatiable curiosity of spectroscopists is only the most obvious suitor of the new device. The substantial density of atoms that seems to have been achieved suggests that it should be possible almost directly to simulate astrophysical processes in, for example, molecular clouds, especially if there is a chance of using the tweezers effect to mix together different materials. But, inevitably, there will soon also be better traps. Pritchard *et al.* have thoughtfully given no fewer than three recipes that will no doubt be quickly followed.

John Maddox

Pattern formation

Descartes and the fruitfly

from Geoffrey North

In recent years remarkable advances have been made in elucidating the developmental biology of the fruitfly *Drosophila melanogaster*, but despite the cloning and intensive analysis of many of the genes controlling morphogenesis, we still do not understand the processes of pattern formation in mechanistic detail. The evidence of a recent meeting*, however, suggests that this goal may at last be in sight. Perhaps the most telling signs of this are the links that are emerging between a complex and rapidly expanding set of data on the genetic control of embryonic segmentation on the one hand, and a theoretical model that can simulate many of the observations on the other.

The meeting also provided what seems likely to be the first identification of bona fide developmental morphogens, which define the anterior-posterior axis of *Drosophila* embryos. A gene implicated in the establishment of dorsal-ventral polarity has turned out to encode a serine protease, which as I shall explain may help to suggest a precise mechanism of pattern formation along this axis. And intriguing evidence was presented that these orthogonal axes might function like a Cartesian coordinate system, their points of intersection specifying different pattern elements at particular positions of the embryo.

Anterior-posterior axis

The polarity of the anterior-posterior axis of an insect is manifested both in the sequence of segments along its body and in the pattern within an individual segment. Various experiments suggest that this originates in localized morphogenic centres laid down in the egg from the onset of development. For example, the egg of the leaf-hopper *Euscelis* contains at its posterior end a ball of symbiotic bacteria, which can be experimentally moved within the egg. Such manipulations induce the development of posterior structures at inappropriate positions along the body axis, presumably because of cytoplasmic determinants carried with the ball of bacteria. Perhaps more dramatically, the ultraviolet irradiation of one or other pole of eggs of the midge *Chironomus* leads to the development of either double-headed or double-abdomen 'monsters' in which the untreated pole has apparently undergone a mirror-image duplication. On the basis of these and other experiments Sander postulated that two gradients,

one from either end, set up the anterior-posterior pattern of developing insects.

It has not, however, been clear how these striking results could be taken much further, so to bring the powerful techniques of classical and molecular genetics to bear on the problem. Christiane Nüsslein-Volhard and collaborators (Max-Planck Institute, Tübingen) have attempted to reproduce the observations with *Drosophila*, which with a relatively small egg is less amenable to physical manipulation. They have succeeded, in a series of experiments involving various combinations of allowing cytoplasm to escape from the embryo through holes punctured at specific positions through the egg-case and transplantation of cytoplasm from one region to another¹. For example, loss of cytoplasm from the anterior end causes loss of head structures and the development at the anterior pole of structures such as the telson normally only found at the posterior pole. If, in addition, posterior cytoplasm is transplanted to the anterior end the result is an embryo consisting of two abdomens arranged in mirror-image symmetry.

The phenotypes produced in these experiments can be reproduced by certain 'maternal-effect' mutations, in which the inability of the mother fly to make a wild-type gene product leads to the mutant embryonic phenotype. Thus *bicoid* mutants lack head structures (Nüsslein-Volhard) and *oskar* mutants lack an abdomen (R. Lehmann, Max-Planck Institute, Tübingen). Cytoplasm from the appropriate end of a wild-type embryo can rescue these mutants if injected into a specific region along their body axis, and in the case of *bicoid* mutants injection at the wrong site causes ectopic head development. The *bicoid-oskar* double mutants are particularly revealing: they lack both head and abdomen, and consist of juxtaposed posterior terminal structures which seem to be a kind of 'ground-state' made in the absence of information to the contrary. If such mutants are injected with wild-type cytoplasm from the anterior or posterior pole it is possible to correct one or other mutation, generating single mutant phenotypes in either normal or inverted orientation, depending on the end into which the cytoplasm is injected.

These results lead to a picture of mutually antagonistic determinants localized at either end of the egg with long-range effects on the nature and polarity of structures formed during development: whether they really generate formal 'gradients' is not yet clear. The *bicoid*

gene, which is located within the Antennapedia complex of homoeotic and segmentation genes, has been cloned, and Nüsslein-Volhard's *in situ* hybridization experiments show that it encodes a transcript localized at the anterior pole of the egg, presumably transported there from the transcriptionally active nurse cells to which the egg is connected.

The *bicoid* gene, therefore, has all the features of a gene encoding a developmental morphogen. The situation for *oskar* appears somewhat more complicated, for together with a small set of other maternal-effect genes, it seems to generate a signal at the posterior end that is transported anteriorly, where it has its effects on zygotic gene expression. In a model put forward by Hans Meinhardt² (Max Planck Institute, Tübingen), a single gradient of maternal information is interpreted by the sequential activation along the anterior-posterior embryonic axis of the zygotic 'gap' genes, mutations of which cause the loss of large contiguous groups of segments. The results with *bicoid* and *oskar* are more consistent with Sander's proposal that there are morphogenetic centres at both poles of the embryo, but the 'gap' phenotype of *oskar* mutants supports the suggestion that the gap genes are the next stage in the developmental hierarchy after the maternally expressed genes.

In Meinhardt's model, the gap genes (which he prefers to call cardinal genes) play a central part in initiating both the periodic pattern of segmentation and, via the homoeotic genes they are postulated to regulate, the sequential pattern of segment specification. Interest in this model has been fuelled by recent results that bear out several of its non-trivial predictions. For example, although different gap mutations delete large, overlapping regions of the embryo, Meinhardt predicted that they would be expressed in non-overlapping regions about half the length of the affected region — and in the case of *Krüppel* and *hunchback* this has turned out to be just the case (H. Jäckle, Friedrich Miescher Institute, Tübingen). Also confirmed has been the prediction that the loss of a given gap gene would cause the expression of the neighbouring gene to spread into its domain (Jäckle). In the model the gross non-autonomy of gap genes results from the central importance of the borders between the domains of gap gene expression in organizing the expression of the next set of genes in the postulated hierarchy, the pair-rule genes (the first to have a periodic pattern of expression): the loss of one gap gene eliminates two borders and hence a region of the embryo twice as long as that in which it is expressed.

Data on the expression of the segmentation genes of *Drosophila* are being obtained at a tremendous rate, and some

*5th EMBO workshop on the molecular and developmental biology of *Drosophila*, Kolymbari, Crete, 21–29 June 1986.

of these have been discussed recently in *News and Views*⁶. The main focus at present is to obtain clues as to how the genes interact by studying the effects of mutations in one gene on the expression of another. With so many genes apparently involved, however, the interpretation of such experiments is unlikely to be simple, and it was notable at the meeting that most of the experimentalists averred that they found Meinhardt's model a helpful framework in which to try to interpret their results. The film Meinhardt showed at the meeting of computer simulations based on his model demonstrated that it can closely simulate the evolving patterns of segmentation gene expression in wild-type and normal embryos that are observed experimentally.

Dorsal-ventral axis

A further set of maternal-effect mutations affects pattern-formation along the dorsal-ventral axis of the *Drosophila* embryo. Recessive mutations in at least 10 genes cause dorsalization of the embryo, and a series of experiments^{7,8} reviewed recently in *News and Views*⁹ has led to a model for the establishment of dorsal-ventral positional information by a morphogenetic product generated by these genes. One of this set, *snake*, has been cloned and sequenced, and encodes a protein with all the features of a serine protease (R. Delotto, University of Geneva). This is an exciting result, for serine proteases are made as inactive precursors activated by specific proteolytic cleavage, and in blood clotting and complement fixation chains of specific serine proteases are arranged in a cascade to amplify an initial signal in such a way as can readily be envisaged to be applied to the more glamorous problem of pattern formation. Both blood clotting and complement fixation generate products tightly localized in space close to the initial signal: the intriguing prospect is that such a signal, probably at the ventral pole, initiates a cascade of reactions which converts it into a set of coordinates along the dorsal-ventral axis. If this is true, other dorsalization genes should turn out to encode serine proteases.

Although the issue has been controversial, it is now generally accepted that during development the anterior-posterior axis of the *Drosophila* embryo is divided into polyclonal units called compartments, defined by the expression of sets of 'specifier' genes that determine their pathways of development. Whether the dorsal-ventral axis is divided into compartments analogous to those along the anterior-posterior axis is unclear. That this may be the case is suggested by recent work on another gene of the Antennapedia complex that has recently been cloned, *zerknüllt* (*zen*)¹⁰. Mutations of this gene affect the differentiation of

A bacterial haemoglobin

from M.F. Perutz

HAEMOGLOBIN was believed to be a protein that evolved in eukaryotes, but a report elsewhere in this issue (Wakabayashi, S., Matsubara, H. & Webster, D.A. *Nature* 322, 481; 1986) describes a bacterial haem protein from *Vitreoscilla* that combines with molecular oxygen and has an amino-acid sequence with features characteristic of the globins.

There are only two amino-acid residues that are common to all the globins: a histidine that forms a covalent link with the haem iron; and a phenylalanine in position 1 of the loop made by helices C and D which wedges the haem into its pocket. All other residues are replaceable, but in 33 specific positions replacements are restricted to non-polar residues (Perutz, M.F., Kendrew, J.C. & Watson, H.C. *J. molec. Biol.* 13, 669; 1965).

The haem protein of *Vitreoscilla* does contain non-polar residues at almost all these positions, and it shows marked homology with the haemoglobin of the leguminous nodules of lupins, whose globin is encoded by a gene in the plant, rather than the symbiotic bacterium *Rhizobium* where the haemoglobin is concentrated. Most haemoglobins contain a histidine in a position distal to the haem, but the α -chain of the opossum is an exception and contains a glutamine in-

stead. The alignment of the *Vitreoscilla* sequence by Wakabayashi *et al.* suggests that it also has a distal glutamine whose amino group could form a hydrogen bond with the bound oxygen, just as the distal histidine does in myoglobin (Phillips, S.E.V. & Schoenborn, B. *Nature* 292, 81; 1981). Homology between *Vitreoscilla* and other globins extends to helices B, C, E, F, G and H; helix D is missing as it is in the α -chains of mammalian haemoglobins, and homology ends at the A-B corner. The first 11 residues show no homology with the A helix, but this helix is unimportant because it is not in contact with the haem.

What is the function of this haemoglobin? The haemoglobin content of *Vitreoscilla* increases almost 50-fold when the oxygen concentration of the growth medium falls below 10 per cent of atmospheric. Apparently, species of this genus live in oxygen-poor environments; possession of the haemoglobin helps them to absorb oxygen for their aerobic metabolism. Is this an example of divergent or convergent evolution or of gene transfer from eukaryotes? Who can guess? We were not there when it happened.

M.F. Perutz is at the Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

dorsally derived embryonic tissues. Analysis of the cloned gene shows that it contains a homoeo box and is expressed in dorsal tissues of developing embryos. The evolution of the pattern of *zen* expression along the dorsal-ventral axis during development is reminiscent of genes such as *Ultrabithorax* that are expressed in a compartment-specific manner along the anterior-posterior axis, suggesting that *zen* may be involved in the specification of a dorsal-ventral compartment.

A Cartesian coordinate frame?

Meinhardt has suggested that the points on the surface on an insect embryo where dorsal-ventral and anterior-posterior compartment boundaries intersect might specify the positions of the imaginal disks that later in development generate the adult cuticle structures. Intriguing evidence that this could be true was reported by W. Gelbert (Harvard). Mutations of the *decapentaplegic* gene of *Drosophila* cause the loss of pattern elements derived from the centres of imaginal disks. Gelbert explained that clonal analysis shows that a normal *decapentaplegic* gene is required only in cells that abut the anterior-posterior compartment boundary of a disk for normal development.

Furthermore, using the cloned gene it has been found that *decapentaplegic* is expressed in a dorsal region of the embryo, later splitting into two stripes running along the anterior-posterior axis on either side of the embryo just where the two lines of imaginal disks later develop. So it could be that these stripes define dorsal-ventral compartment boundaries, and the spatial cue for disk formation is where they cross anterior-posterior compartment boundaries. I am sure that the founder of analytical geometry, René Descartes, would be pleased by this elegant application of his system of spatial coordinates to biological pattern formation. □

1. North, G. *Nature News and Views* 311, 214 (1984).
2. Sander, K. *Adv. Insect Physiol.* 12, 125 (1976).
3. Frohnhofer, H.G., Lehmann, R. & Nüsslein-Volhard, C. *J. Embryol. exp. Morph.* (in the press).
4. Lehmann, R. & Nüsslein-Volhard, C. *Cell* (in the press).
5. Meinhardt, H. *J. Cell Sci. Suppl.* 4, 357 (1986).
6. Coulter, D. & Wieschaus, E. *Nature News and Views* 321, 472 (1986).
7. Anderson, K.V., Jurgens, G. & Nüsslein-Volhard, C. *Cell* 42, 779 (1985).
8. Anderson, K.V., Boklén, L. & Nüsslein-Volhard, C. *Cell* 42, 791 (1985).
9. Woodland, H. & Jones, E. *Nature News and Views* 319, 261 (1986).
10. Doyle, H.J., Harding, K., Hoey, T. & Levine, M. *Nature* (in the press).

Geoffrey North is Deputy Biology Editor of *Nature*.

Neural models

Networks for fun and profit

from James A. Anderson

KNOWING how the brain works is intrinsically interesting and could be practically useful. There have been many attempts to make artificial systems compute the way we seem to. Recently, there has been a renaissance of interest in what have been variously called neural models, brain models, neural networks and connectionist models. Two recent papers^{1,2} provide examples of how some important practical problems turn out to be well suited to computations based on neural networks, and applications using these networks may appear in the near future.

Connectionist systems were originally proposed as models for brain organization and for psychological functions such as association, concept formation and word recognition. But there has been a change in the past year — an adequate brain model must be able to compute something. It seems that even at their current state of development, the things that connectionist models compute effectively are now of interest to many scientists and engineers with practical problems.

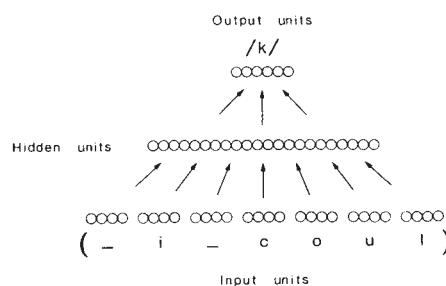
There are several variants of connectionist models, but all have similar structure. They assume that there are a very large number of simple processing elements arranged in parallel arrays. Processing elements can either be simple Boolean devices with two allowable states (for example, on or off) or they can have continuously graded activity. Elements in the same group can be coupled together or can be joined to elements in other groups by connections with modifiable strengths. This interconnected system is largely based on the architecture of the mammalian cerebral cortex, with varying degrees of oversimplification: the properties of the elements are inspired by the properties of neurones and those of their modifiable connections by synapses.

There are two distinct aspects to the functioning of a connectionist system. First, the strengths of the interconnections must be specified. One way this can be done is by having the network learn a set of training activity patterns so it forms the appropriate connection strengths based on a learning rule. Most learning rules are variants of an idea proposed by Donald Hebb in 1949, that a connection strength is changed based on the simultaneous occurrence of activity on both sides of the connection. Another way is by simply knowing what the strengths should be to compute what it is you want to know. This is possible for some well-defined problems, and has the great advantage that a slow and sometimes unreliable learning

process can be bypassed.

Second, when the connections are formed, the network must be provoked to perform the desired computation. An input pattern of activity is provided to start the system, then the initial pattern, through the interconnections, modifies itself and affects the activity of other connected elements to form a new pattern. This process may repeat for a number of steps until the network stabilizes, or it may run continuously. If the system works properly, the new patterns contain the results of the computation.

One example of a new application of networks is NETalk, a program which converts text into spoken English (reads) developed by Terrance Sejnowski and Charles Rosenberg¹. Reading printed text



Schematic drawing of the network architecture for NETalk. Input units are shown on the bottom of the pyramid, with 7 groups of 29 units in each group. A 29-unit group corresponds to a single letter, giving rise to a 7-letter window into the text. Each hidden unit in the intermediate layer receives inputs from all the input layers in the bottom layer, and in turn sends its output to all 26 units in the output layer. An example of an input string of letters is shown below the input groups, and the correct output phoneme for the middle letter is shown above the output layer. For 80 hidden units, which were used for the corpus of continuous informal speech, there was a total of 309 units and 18,629 weights in the network. (Figure and caption from ref. 1.)

is typical of the rule-based systems used by humans. There are important rules and regularities in English spelling, but there are also many exceptions to the rules. Sejnowski and Rosenberg used samples of English text with the conversion of that text into phonetic symbols. As they then knew the input strings of characters and output strings of phonemes for many examples, they could modify connection strengths using an effective learning algorithm called back propagation, developed by David Rumelhart, Geoffrey Hinton and Ron Williams², and independently by David Parker³ and by Yann Le Cun⁴.

This algorithm is able to learn effectively when there are several layers of modifiable connections between input and output. It is an error-correction technique that sends corrections 'backwards' from the output (where we can measure the error) to earlier layers with the appropriate strengths. NETalk looks at seven input characters at once to generate an output phoneme, as pronunciation of a particular letter is strongly dependent on context. There are three layers of processing elements: an input layer, an output layer and a layer of 'hidden units' between input and output. After about a dozen hours of computer time on a VAX superminicomputer, the program was 'speaking' correctly more than 90 per cent of the words encountered in the input text. When a speech synthesizer was connected to the output, it generated quite understandable speech with occasional, usually minor, mispronunciations.

This simulation is much more than a curiosity. Sejnowski modelled the name of the simulation, NETalk, after a commercial product of the Digital Equipment Corporation called DECtalk, a complex system developed over several years which speaks the words and characters appearing on a computer terminal. Sejnowski and Rosenberg, starting from zero, took less than three months to develop a functioning system that can do some of the same things.

A second example is the recent work of John Hopfield and David Tank⁵. Hopfield had written two influential papers^{6,7} that pointed out some similarities between neural network computations and some problems in solid-state physics, for example, spin glasses, with a similar mathematical structure. He showed that the essence of the dynamics of a network computation in many cases is that the network is trying to put the state of the system in a minimum of what can be interpreted as an energy function in a physical system. This means that many of the powerful analytical tools available in physics can be carried over bodily to neural networks and might help towards understanding brain function.

The new paper by Hopfield and Tank⁵ shows how a neural network can be applied to a practical problem that seems remote from brain function. They discuss the use of network model to estimate the answer to the 'travelling salesman' problem, the best-known example of a set of problems called *np*-complete by computer scientists. Like many good problems it is simple: a travelling salesman is given a number of cities to visit on his route. What sequence of cities gives the shortest total travel distance? To be certain the shortest route has been found, it is necessary to compute the distance for all possible routes. As the number of cities increases, the number of combinations to check

grows exponentially, and it rapidly becomes impractical to find the best solution. This problem is related to many optimization problems such as minimizing costs and maximizing efficiency.

Hopfield and Tank² show that a suitably designed neural network can compute a good estimate of the best solution. When checked, it is not the best solution but it is quite close to the best and can be computed quickly. Their technique is to arrange the system so that only allowable solutions to the problem can occur. For example, if the city is visited eighth on the circuit, then no other city can also be visited eighth. When one city is chosen for that position on the trip, it inhibits all the other candidates for the eighth position. No city can be visited twice, so if a city is chosen for one position on the route, it inhibits its potential appearances in other positions on the route. Hopfield and Tank constructed the connection strengths and the energy function so that as the state of the network evolves, the states corresponding to routes with small total distance grow at the expense of other routes. If the initial state of the system is random, the final state of the system corresponds to a route with a small total distance and is an allowable route. If the computation is repeated a few times with different initial states, the best solution is close to the actual solution of the problem. For many practical applications, this is good enough to be satisfactory.

But special algorithms for this problem are able to get better and faster estimates, and DECTalk works. So why bother with NETalk and its potential descendants? One reason is that these two quite different problems are solved with techniques that are basically that same: interconnected sets of simple elements coupled with modifiable connections. Special-purpose hardware built to make one compute quickly will make the other compute quickly as well. Also, the algorithms for a network computation are highly parallel, that is, all the elements are computing their outputs simultaneously. Network computations are therefore well suited for parallel computers. For example, optical hardware, with its high speed, low precision, parallelism and high interconnectivity, is an excellent match for the requirements of network computations. Several groups are now building special-

purpose parallel hardware specifically for networks.

Once the hardware becomes available, it can be used for a wide range of problems, from optimization to learning rules for natural language, with only minor changes. Connectionist models may not give the best or fastest answer to any particular problem but will come up with a

pretty good answer quickly. It is this potential flexibility and speed that makes practical applications so attractive. The brain may show similar computational flexibility and power for similar reasons. □

James A. Anderson is in the Department of Psychology, Center for Neural Science and Center for Cognitive Science at Brown University, Providence, Rhode Island 02912, USA.

Physiology

Chloride ions and cystic fibrosis

from Maynard Case

CYSTIC fibrosis, a lethal inheritable disease, seems to involve defective transport of chloride ions across epithelia. Such a defect could be caused by an alteration in the number, characteristics or regulation of Cl⁻ channels in the plasma membrane of epithelial cells. Three recent articles¹⁻³, one of which appears on page 467 of this issue, suggest it is the regulation of Cl⁻-channel activity that is at fault.

The conventional electrophysiological studies on sweat gland ducts¹ and airway epithelia² (which are both affected in cystic fibrosis) point conclusively towards the involvement of decreased Cl⁻ transport in the pathophysiology of this disease. The nature of this transport defect has now been studied independently by Welsh and Liedtke^{1,2} and by Frizzell *et al.*³ using patch-clamp techniques.

The two groups have studied primary cultures of tracheal epithelial cells from normal subjects and cystic fibrosis patients and have recorded both from intact cells and excised, inside-out membrane patches. They both describe the presence of Cl⁻-selective channels on the apical membrane of these epithelial cells but under similar experimental conditions, Welsh and Liedtke observe only one type of channel whereas Frizzell *et al.* detect two channels. Also, the channel described by Welsh and Liedtke seems to be insensitive to Ca²⁺ but those described by Frizzell *et al.* are both Ca²⁺-sensitive.

Despite these apparent and important differences in the characteristics of the channels, which might perhaps reflect differences in tissue culture technique, the major findings of the two studies are similar. First, during recording from intact cells, Cl⁻-channel activity can be elicited by adrenaline or cyclic AMP in normal but not cystic fibrosis cells; and second, in excised patches, the biophysical properties and kinetic behaviour of the channels are identical in both cell types. This suggests that the problem in cystic fibrosis relates not to the presence or absence of the Cl⁻ conductance but to the regulation of this conductance by a cyclic AMP-dependent process.

Although both studies describe differ-

ences between normal and diseased cells only about half the intact normal cells respond to stimulation by opening Cl⁻-channels whereas all excised patches from normal and diseased cells show activity. This may indicate active inhibition of the Cl⁻ channel so that excising the patch or stimulating the cell removes some sort of brake. If so, perhaps the brake-release mechanism is at fault in cystic fibrosis. Alternatively, there could be two populations of epithelial cells in normal subjects, but only one in cystic fibrosis patients.

These articles refocus our attention on the pathway generating cyclic AMP, but as adrenaline causes the normal accumulation of cyclic AMP in cultured airway cells from cystic fibrosis patients, presumably the defect lies downstream from cyclic AMP generation itself. If a cyclic AMP-dependent pathway is involved in the disease, this may help to explain the effects of the disease on the pancreas — another prime target.

Secretion in the pancreas occurs across both acinar and ductal epithelia. Whereas the former is controlled by Ca²⁺-mobilizing agonists (cholecystokinin and acetylcholine), secretion across the ducts is controlled by secretin, which seems to act exclusively via cyclic AMP. Hence these studies on airway epithelia support the hypothesis that pancreatic ductal secretion is faulty in the disease. However, as cystic fibrosis is thought to be associated with defective Cl⁻ secretion, it has always been a problem to explain why an epithelium lining the pancreatic ducts that secretes HCO₃⁻, rather than Cl⁻, should be affected (although this HCO₃⁻ secretion is Cl⁻-dependent). If the fault lies in the regulation of the anion channel rather than the channel itself, as this recent work suggests, this objection is overcome. □

1. Welsh, M.J. & Liedtke, C.M. *Nature* **322**, 467 (1986).

2. Welsh, M.J. *Science* **232**, 1648 (1986).

3. Frizzell, R.A., Reckemmer, G. & Shoemaker, R.L. *Science* **233**, 558 (1986).

4. Quinton, P.M. *Nature* **301**, 421 (1983).

5. Knowles, M.R. *Science* **221**, 1067 (1983).

Maynard Case is in the Department of Physiology, University of Manchester, Manchester M13 9PT, UK.

1. Sejnowski, T.J. & Rosenberg, C.R. *The Johns Hopkins University Electrical Engineering and Computer Science Technical Report JHU/EECS-86/01* (Baltimore, Maryland, 1986).

2. Hopfield, J.J. & Tank, D.W. *Biol. Cybernet.* **5**, 141 (1985).

3. Rumelhart, D.E., Hinton, G.E. & Williams, R.J. in *Parallel, Distributed Processing: Explorations in the Microstructure of Cognition* (eds Rumelhart, D.E. & McClelland, J.L.) (MIT, Cambridge, in the press).

4. Parker, D. *MIT Tech. Rep.* TR-47 (1985).

5. Yann Le Cun in *Disordered Systems and Biological Organization* (eds Bienenstock, E., Fogelman Soulie, F. & Weisbuch, G.) (Springer, Berlin, in the press).

6. Hopfield, J.J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2554 (1982).

7. Hopfield, J.J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3088 (1984).

Population biology

Genetic management in zoos

from Andrew F. Read and Paul H. Harvey

ANIMAL populations in zoos tend to be small and, therefore, particularly susceptible to the loss of genetic variability. Such losses can have harmful effects on the fitness of individuals as well as reducing the evolutionary potential of populations. Interest in the genetic management of zoo populations has recently taken on a fresh urgency² because of the increasing realization that captive breeding programmes are essential to prevent many species from becoming extinct. A new 13-chapter report, written by about 40 zoo managers, geneticists and population biologists³, introduces rather than solves many relevant problems, but should provide an important framework for future research.

The increasing size of the human population is causing large-scale degradation of tropical and temperate ecosystems that support most of the world's wildlife. If the human population achieves zero population growth in about 100 years, as some demographic studies forecast^{4,5}, and then begins to decline, it would take several centuries before viable populations of many animal species could be restored in the wild. Given such dismal prospects for the world's fauna, zoos must be prepared to act as stewards or as 'arks', to use the metaphor of Soulé *et al.*³. These authors point out that technical improvements may eventually provide other ways of maintaining populations (for example, by regenerating organisms from preserved zygotes). But in the meantime, they argue, about 2,000 species of terrestrial vertebrates, including 160 primates, 100 large carnivores, 100 artiodactyls, 800 birds and several hundred amphibians and reptiles, may have to be captively bred to save them from extinction.

The preservation of genetic diversity is becoming a primary goal of many captive breeding programmes. Juvenile mortality in captive species has been linked to the loss of genetic diversity after inbreeding⁶. Similarly, the low level of heterozygosity in South African cheetahs *Acinonyx jubatus jubatus* may underlie the great difficulty these animals have breeding in captivity, their high degree of juvenile mortality and their extraordinary sensitivity to pathological viruses⁷. Loss of genetic diversity could also limit the potential of a population to adapt to new environments when reintroduced to the wild.

If maintenance of genetic variation is to be a central aim of management, how should we measure it? Electrophoretically detected protein variation has been used to estimate the overall extent of genetic variation within populations. For example,

such tests identified the low level of heterozygosity in cheetahs⁸. Allozyme surveys can be used to distinguish outbred populations from genetically homogeneous inbred populations, but heterozygosity at a few protein-coding loci is not necessarily a good estimate of the overall genomic heterozygosity of an individual. For example, the average number of loci examined in routine electrophoretic surveys is about 20. As Hedrick and co-authors point out³, if the expectation of heterozygosity is the same at all loci; if the loci segregate independently; and if the 20 loci are examined from a pool of 1,000, the correlation between heterozygosity for those loci and overall heterozygosity is only 0.14.

Furthermore, heterozygosity may be an

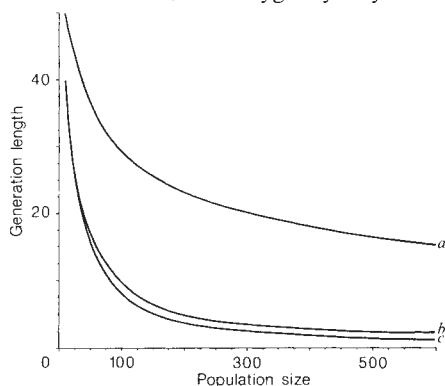


Fig. 1 Combinations of target population sizes and generation lengths required for a species to retain 90 per cent of the genetic variation of the source population for 200 years, given an exponential growth rate of 1.0 and founder population sizes of 10 (a), 20 (b) and the target number of individuals (c)³. Species with long generation times can be maintained with fewer animals, and the size needed can be decreased dramatically by increasing the number of founders used to initiate the captive population.

insufficient measure of genetic variation. Allendorf argues that useful allelic diversity is largely caused by rare alleles that are likely to be lost early in a captive breeding programme, with only a minimal effect on heterozygosity³. Although heterozygosity is a good measure of the ability of a population to respond to selection in the short term (the first few captive generations), the number of alleles may be a more important indicator of the potential long-term response under continued selection. Thus, both high heterozygosity and high allelic diversity are desirable in species under consideration for reintroduction programmes. The problem, then, is to determine the number and distribution of alleles. Electrophoretic

techniques provide somewhat conservative estimates, whereas DNA sequencing is more accurate but more costly.

How many individuals of each endangered species should be kept in zoos? This depends on the amount of genetic variation to be maintained in the captive population over a specified amount of time. Soulé and co-authors set themselves the theoretical problem of keeping a captive population which, after 200 years, retains about 90 per cent of the genetic variation found in the original wild population³. This, they think, is the intuitive threshold between a tolerable loss of heterozygosity and potentially damaging homozygosity that provides enough time for the biotechnology to develop that will be necessary to maintain and regenerate species from preserved eggs, zygotes or tissues. Their model assumes random breeding in groups, equal sex ratios, non-overlapping generations and Poisson-distributed family sizes. The target captive population size depends on generation time, population growth rate and the number of interbreeding founders. The combinations of population sizes and generation lengths required for species with an exponential growth rate of 1.0 are shown in Fig. 1. Given that a major aim is to use resources as efficiently as possible, zoo managers are faced with the prospect of minimizing the number of individuals of each species held, while at the same time attempting to retain genetic variability within the captive population.

For a fixed number of founders, species with longer generation times require smaller target captive population sizes, because under the assumptions of the model there is loss of alleles only during reproduction. But the required target population sizes can be very sensitive to the number of individuals taken from the wild. For the values shown in Fig. 1, a large decrease in the target population size can be achieved by increasing the number of founders, at least up to 25 animals. (Fewer founders are needed in species with high population growth rates but, even with the highest rates of growth, the founder population should still be above six, otherwise more than 10 per cent of the genetic variability will be lost in the first generation.) Ideally then, captive propagation plans should aim to start with more than 20 founders (although the actual number depends on the rate of growth and generation time of the species in captivity), increase to the eventual population size in few generations (so that the effects of genetic drift are minimized and to ensure survival of the species in captivity) and then, if possible, artificially delay reproduction. Soulé and co-authors consider that breeding programmes should aim to maintain an effective population size of 200–300 individuals for more species.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The number of founders and reproductive rates of several endangered species currently in captive breeding programmes falls well short of those guidelines. The population of Przewalski's horse (*Equus caballus przewalskii*; Fig. 2), for example, has a growth rate of 0.1 and is derived from only 13 founders. According to Soule's model there is no target population size which will allow the retention of more than 90 per cent of the original genetic variability.

Should populations be divided? Subdivision helps to protect populations from total extinction by epidemics or local disasters, and for this reason alone it is desirable to have more than one captive group of any endangered species. Foose and co-authors also point out that, if at least one breeding individual per generation is transferred between groups and there is no differential mortality according to genotype, the entire captive population is nearly equivalent to a single interbreeding unit⁹.

Which animals should be chosen for captive breeding programmes? This is by no means a trivial question. The choice of individuals for reintroduction to the wild must take into account the extent to which they are likely to be adapted to local conditions. For example, the Tatra Mountain ibex *Capra ibex ibex* became extinct in Czechoslovakia but was successfully reintroduced from Austria. Subsequently, *C. i. aegagrus* from Turkey and *C. i. nubiana* from Sinai were added to the Czech herd. The resulting hybrids rutted in the early

autumn, rather than in the winter as the local ibex did, and their young were born in February, the coldest month of the year. As a result the whole Czech population became extinct¹⁰.

Major problems facing captive breeding programmes include relaxation of natural selection and inadvertent selection for domestication. For example, selection for ease of handling may result in loss of predator escape behaviour. Frankham and co-authors conclude that selection in captivity can be minimized by adopting breeding plans designed to maximize genetic diversity¹. As well as maximizing the number of founders, rapidly expanding and then subdividing the population, they suggest equalizing the genetic contribution of each of the founders to subsequent generations by manipulating family sizes. The only active selection they advocate is to cull animals with specific deleterious traits and outliers in morphological or fitness measures.

Any other forms of artificial selection must be justified. Selection for captive breeding success is clearly appropriate for rare species not yet capable of self-sustaining reproduction in captivity. In some cases it may be desirable to select against certain genotypes. Captive populations of Przewalski's horse, for example, have had genetic inputs from 12 wild-caught Przewalski's horses and one Mongolian domestic mare. The historical distribution of the breeding lines has resulted in two-thirds of the current population having genetic inputs from this

mare. An interim breeding plan aimed at purifying the captive population by allowing matings only between animals lacking genetic input from the mare has begun¹¹. This selection will, of course, further reduce the genetic diversity of the species. In the case of the Speke's gazelle *Gazella spekei*, selection for inbreeding tolerance was applied. This was inevitable because the captive population was facing serious inbreeding problems, had very few founders and no opportunity for input from the wild².

Hedrik and co-authors add a note of caution³. Successful reintroduction, they say, is more likely when genetic variation in a captive population is maintained in the form found in natural populations. The maintenance of the breeding and social structure found in the wild may be more important for a species than a programme designed only to retain maximum genetic diversity.

Several simulation models have been developed to assess the interacting effects of such competing management goals. These models assign a genotype to each founder and then, using pedigree data and demographic characteristics of populations, subject the 'colony' to various management plans. One such computer experiment, reported by Dyke and co-authors⁴, uses data from a real colony of captive rhesus monkeys to test the idea that removing some individuals from a single age group (usually early in life) and removing all older breeders with low reproductive rates would increase the



GIANT tube worms (*Lamellibrachia* sp.) have been found in Japan's deep-sea back yard, 1 km from the east coast of the Izu Peninsula, at a depth of 1,160 m. The worms, 60–70 cm in length, occur alongside giant clam (*Calyptogena soyoe*) colonies, the largest of which spans 200 by 80 m. Methane levels reach 10,000 μ l per litre a few centimetres above this 'vent' community which is aligned along a major fault associated with subduction of the Philippine plate beneath Japan. Photograph taken in May/June 1986 during a survey by the Japan Marine Science and Technology Center and the Ocean Research Institute, University of Tokyo.

David Swinbanks

1. Ralls, K. *Biologist* **31**, 99 (1984).
2. Foose, T. in *Genetics and Conservation: a Reference for Managing Wild Animal and Plant Populations* (eds Schonewald-Cox, C.M., Chambers, S.M., MacBryde, B. & Thomas, W.L.) 374 (Cummings, Menlo Park, California, 1983).
3. Ralls, K. & Ballou, J. (eds) *Zoo Biol.* **5** (1986).
4. United Nations *Pop. Bull.* **14** (1982).
5. World Bank *World Development Report* (World Bank, Washington DC, 1984).
6. Ralls, K. & Ballou, J. in *Genetics and Conservation: a Reference for Managing Wild Animal and Plant Populations* (eds Schonewald-Cox, C.M., Chambers, S.M., MacBryde, B. & Thomas, W.L.) 164 (Cummings, Menlo Park, California, 1983).
7. O'Brien, S.J. *et al. Science* **227**, 1428 (1985).
8. O'Brien, S.J., Wildt, D.E., Goldman, D., Merrill, C.R. & Bush, M. *Science* **221**, 459 (1983).
9. Crow, J.F. & Kimura, M. *An Introduction to Population Genetics Theory* (Harper & Row, New York, 1970).
10. Greig, J.C. *S. Afr. J. Wildl. Res.* **9**, 57 (1979).
11. Ryder, O.A. & Wedemeyer, E.A. *Biol. Conservation* **22**, 259 (1982).
12. Templeton, A.R. & Read, B. in *Genetics & Conservation: a Reference for Managing Wild Animal and Plant Populations* (eds Schonewald-Cox, C.M., Chambers, S.M., MacBryde, B. & Thomas, W.L.) 241 (Cummings, Menlo Park, California 1983).

productivity above that obtained when individuals were removed at random. Productivity increases with minimal effects on the genetic variability of the population. The implementation of increased population growth rates need not, therefore, add significantly to the problems of genetic management.

Ultimately, the biological problems so carefully brought together in this volume may be dwarfed by financial and political

constraints. At present, zoos can sustain around 925 species of terrestrial vertebrates with captive population sizes of more than 250, far short of the 2,000 species that may need to be accommodated. This time around, the ark may be very crowded. □

Andrew Read and Paul Harvey are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Archaeology

A new chronology for the French Mousterian period

from Paul Mellars

THE classic sequence of 'Mousterian' industries recorded within the cave and rock shelters of southwestern France represents an important record of Neanderthal man, but the dating has been controversial. Dating of the Le Moustier site by thermoluminescence techniques, reported on page 452 of this issue, provides one of the first well-documented points of reference for the chronology of the sequence and therefore will require reanalysis of climatic and human changes during this period.

The time range of the Mousterian (broadly from 40,000 to 115,000 years before present (BP)) — coinciding with the earlier part of the last glaciation) puts it effectively beyond the range of dating by means of conventional radiocarbon techniques, and even the refinements of the new accelerator mass spectrometer techniques of ^{14}C measurement are unlikely to extend for more than 10–20,000 years into the most recent part of the period².

All attempts to construct a detailed in-

ternal chronology for the Mousterian have had to rely either on complex and controversial correlations between the geological and palaeo-environmental sequences recorded in individual sites^{3,5}, or on equally contro-versial correlations between the character of the archaeological material recovered from the sites⁶. But thermoluminescence dating of burnt flint artefacts promises to put the dating of Mousterian sites on to a secure scientific footing.

No dating method is perfect and thermoluminescence dating of burnt flints depends on a rather complex series of calculations — and in some cases assumptions — involving not only the measurement of the thermoluminescence activity of the samples but also the environmental history of the sites and levels dated⁷.

Fortunately there are several internal cross checks on the dates obtained for the Le Moustier sequence. The results of the thermoluminescence dating can be compared directly with the stratigraphic

sequence of the samples, and the dates obtained for the uppermost part of the sequence (containing an early Upper Palaeolithic industry) can be compared with the known age of these levels as documented by radiocarbon dating. In the latter case, thermoluminescence results turn out to be slightly older than would have been expected from the radiocarbon evidence, but only marginally so, when allowance is made for the relatively large standard deviations attached to the thermoluminescence

measurements. The consistency and internal coherence of the dates obtained for the different levels provides strong corroborative evidence for the absolute chronology proposed for the Le Moustier sequence as a whole.

The present chronological structure of the French Mousterian hinges almost entirely on the correlation between the geological and archaeological successions in two key sites — Le Moustier on the one hand, and Combe Grenal (a major rock-shelter site in the Dordogne valley) on the other. Previous interpretations of the geological evidence have assumed that the sequences in both sites span broadly the same range of time, and provide essentially parallel records of occupation throughout at least the greater part of the Mousterian sequence (that is, the 'Wurm I' and 'Wurm II' climatic phases of the current literature)^{3,5}. Since the archaeological material recovered from the two sites is strikingly different, this interpretation carries with it the important implication that distinct industrial variants of the Mousterian were manufactured synchronously in southwestern France throughout most if not all the Mousterian period^{3,5}.

The new thermoluminescence dating of the Le Moustier sequence¹ argues against this interpretation in several ways. In the first place, the earlier geological studies of Laville had postulated a substantial depositional hiatus within the uppermost part of the Le Moustier sequence (between the final Mousterian and earliest Upper Palaeolithic levels)^{4,5} which is now seen to be untenable. More significantly, the new thermoluminescence dates show that the entire sequence of dated levels at Le Moustier (layers G to K) spans a period of only about 15,000 years, which is in sharp conflict with the earlier suggestion that these levels span a period of at least 30–35,000 years^{5,8}. Clearly, the present geological interpretations of the Le Moustier sequence will need to be substantially revised.

The long sequence of Mousterian levels recorded in the site of Combe Grenal^{9,10}, on the other hand, demonstrably spans a very much longer period. The basal levels of the sequence can be reliably attributed on both geological and archaeological grounds to the final part of the penultimate ('Riss') glaciation, while the overlying levels reveal a detailed pattern of climatic fluctuations which parallels almost exactly that recorded for the earliest stages of the last glaciation in recent studies of deep-sea cores (that is, stages 5 and 4 of the ocean-core sequence; see Fig. 1)^{8,9,11–13}. The geological evidence is reinforced by the archaeological record, which shows that classic 'Mousterian of Acheulian Tradition' industries — of the kind which occupy the greater part of the archaeological sequence at Le Moustier — occur only in the uppermost

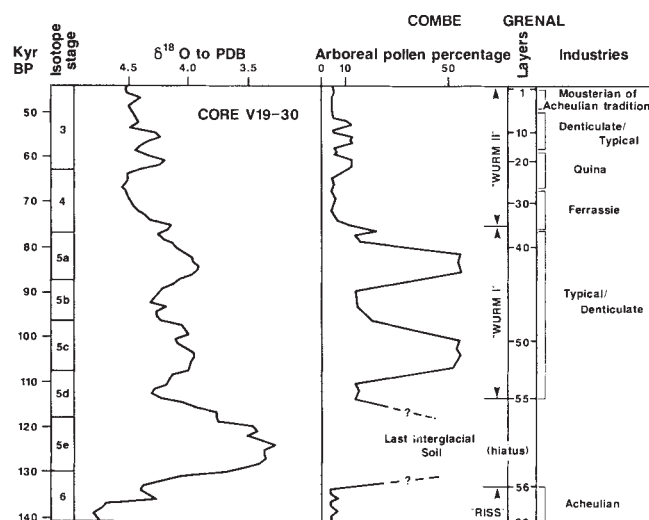


Fig. 1 Suggested correlation between the climatic sequence at Combe Grenal and the sequence of oxygen isotope stages recorded in deep-sea core V1930 (from refs 8, 9, 13 and N. J. Shackleton, personal communication). For details of archaeological sequences, see Fig. 2.

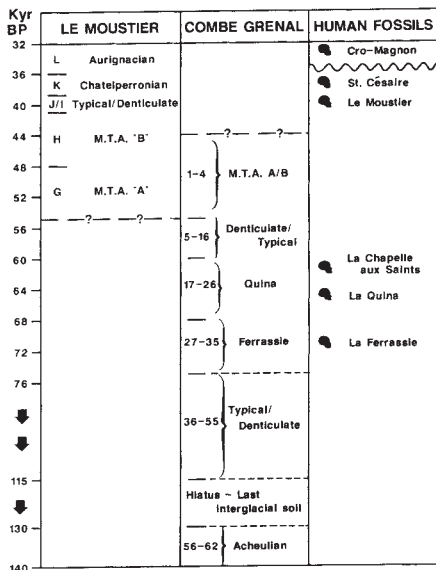


Fig. 2 Correlation between the archaeological sequences at Le Moustier and Combe Grenal implied by the new thermoluminescence dates for Le Moustier. The inferred chronological distribution of the principal Neanderthal remains from south-west France is shown in the right hand column (from refs 5, 10, 14, 15, 17). The sequence of Ferrassie and Quina Mousterian horizons at Combe Grenal is interrupted by a single level of 'Denticulate' Mousterian in layer 20, and three levels of either 'Typical' or 'attenuated Ferrassie' industries in layers 28-30; see ref. 17). M.T.A.; Mousterian of Acheulian Tradition.

occupation levels at Combe Grenal (see Fig. 2). The geological, archaeological and thermoluminescence evidence now converge in suggesting that the two sequences at Le Moustier and Combe Grenal represent essentially successive episodes in the total sequence of Mousterian occupation in south-west France, and together provide a largely complete record of human occupation in the area over a period of at least 70–80,000 years (Fig. 2).

This new chronology has radical implications for almost all aspects of our current understanding of the Mousterian period. In the first place it is now clear that any attempt to force the total pattern of climatic fluctuation during the earlier part of the last glaciation into a simple, two-fold scheme (that is, the 'Wurm I' and 'Wurm II' episodes referred to above) must be seen as a serious oversimplification of a much longer and more complex climatic succession. This in turn has direct implications for the interpretation of the archaeological record. Now that the archaeological sequences at Le Moustier and Combe Grenal can be seen to be successive rather than parallel, there can no longer be any objection to the hypothesis of a basic chronological sequence with three of the major industrial variants of the Mousterian within south-west France (that is, Ferrassie, followed by Quina, followed by Mousterian of Acheulian Tradition) — as the sequence recorded at

Adrian Edmund Gill (1937–1986)

ADRIAN GILL, who died in hospital on 19 April after a sudden illness, will be remembered for his insight, and the almost uncanny ability he possessed to ask apparently very simple questions about ocean and atmosphere.

During the 1970s Gill — who took his doctorate under G.K. Batchelor at the University of Cambridge — began work on a seminal textbook (*Atmosphere–Ocean Dynamics*, published in 1982). The wide reading of both current and historical research that this involved led to the publication of many of the elegantly simple papers for which he will be remembered. His subjects included stability theory, thermal convection, internal and edge waves, and interactions between atmosphere and ocean. The lucidity of his book encapsulated his thinking for the current generation of students and researchers, and rapidly became a best-seller.

Simplicity was the key to the man and his thinking: Gill was a superb mathematician, yet one whose ability was seldom demonstrated, as this would confuse the message that he was anxious to get across not only to his theoretical colleagues but also to seagoing observationalists world-

wide. He was perpetually enthusiastic about discovering more of the physical world in which he lived. Those in Cambridge during his time there will recall the stub of pencil which lived in his jacket and which would always be in evidence at tea-time when some fluid dynamical problem would be hotly debated by those around him; or the essay on anomalous tidal flows in Greece that resulted from a family holiday in that country.

Gill travelled widely in the cause of science, and served on innumerable international working groups and committees. His concern with the world climate led to his becoming a founder member of the Committee for Climate Change and the Ocean, and to the conviction in the last years of his life that a model which coupled the global atmospheric circulation to the tropical ocean was the minimum necessary to understand short-term climatic variations such as El Niño. At the time of his death he was serving as the chairman of the Tropical Ocean–Global Atmosphere programme, one of the main components of the World Climate Research Programme, and had just been elected to the Royal Society.

Peter D. Killworth

Combe Grenal had already suggested (see Fig. 2)^{3,5,6}. The same archaeological sequence is repeated in at least twelve other sites in the same region⁶.

Probably the most significant implications of the new dating however, are for the relative and absolute chronology of the various Neanderthal remains recovered from sites in south-west France. The famous Neanderthal skeleton from Le Moustier itself is known to have come from the uppermost part of the Mousterian sequence¹³ and therefore dated at about 40–45,000 years BP. By contrast, most of the other well-preserved Neanderthal finds are associated with archaeological material which is quite different from that recorded anywhere within the Le Moustier sequence but which compares closely with that recorded in the sequence of Ferrassie and Quina Mousterian levels at Combe Grenal^{14,15}. It would now appear that all of these finds — including the famous 'classic' Neanderthal skeletons from La Ferrassie, La Quina and La Chapelle aux Saints — most probably date from the period of extremely harsh, glacial conditions represented by stage 4 of the ocean core sequence. If so, these finds must be dated at about 60–75,000 years BP. This is at least 15,000 years earlier than the Le Moustier skeleton, and probably 30,000 years earlier than the oldest finds of *Homo sapiens sapiens* forms from Cro Magnon and elsewhere (see Fig. 2).

Exactly what implications this dating may have for current views on the re-

lationships between the various Neanderthal finds and their ultimate relationships with the Cro Magnon forms remains to be seen¹⁶. It is already clear, however, that our whole perception of the chronology of human developments during the earlier part of the last glaciation is beginning to change and it may well be that further application of absolute dating techniques will reveal further surprises.

- Valladas, H., Geneste, J.M., Joron, J.L. & Chadelle, J.P. *Nature* **322**, 452 (1986).
- Hedges, R.E.M. & Gowlett, J.A.J. *Sci. Am.* **254**, 100 (1986).
- Laville, H. *World Archaeol.* **4**, 321 (1973).
- Laville, H. *Climatologie et Chronologie du Paléolithique en Périgord. Etude Sédimentologique de Dépôts en Grottes et sous Abris* (Etudes Quaternaires. Université de Provence, Marseilles, 1975).
- Laville, H., Rigaud, J.-P. & Sackett, J.R. *Rock Shelters of the Périgord: Geological Stratigraphy and Archaeological Succession* (Academic, New York, 1980).
- Mellars, P.A. *Proc. Prehistoric Soc.* **35**, 134 (1969).
- Aitken, M.J. *Thermoluminescence Dating* (Academic, New York, 1985).
- Laville, H., et al. *Bulletin de l'Institut de Géologie du Bassin de l'Aquitaine* (Bordeaux) **34**, 219 (1983).
- Bordes, F., Laville, H. & Paquereau, M.-M. *Actes de la Société Linnéenne de Bordeaux* **103B**, 1–19 (1966).
- Bordes, F. *A Tale of Two Caves* (Harper & Row, New York, 1972).
- Shackleton, N.J. & Opdyke, N.D. *Quat. Res.* **3**, 39 (1973).
- Shackleton, N.J. *Phil. Trans. R. Soc. B* **280**, 169 (1977).
- Shackleton, N.J., Hall, M.A., Line, J. & Chang, S. *Nature* **306**, 319 (1983).
- Bordes, F. *L'Anthropologie* **63**, 154 (1959).
- Vandermeersch, B. *Annales de Paléontologie (Vertébrés)* **51**, 69 (1965).
- Stringer, C.B., Hublin, J.J. & Vandermeersch, B. in *The Origins of Modern Humans: a World Survey of the Fossil Evidence* (eds Smith, F.H. & Spencer, F.) **51** (Liss, New York, 1984).
- Bordes, F. *Bulletin de la Société Préhistorique Française* **52**, 426 (1955).

Paul Mellars is in the Department of Archaeology, Downing Street, Cambridge, CB2 3DZ, UK.

Speciation and our own species

SIR—It is dangerous to build a methodological artefact into dogma. Students of the modern biota are constrained to study those phenomena which are amenable to their methods. Allopatric speciation and other processes of phyletic branching can be studied using modern organisms. Speciation without branching (phyletic speciation) perhaps cannot, so it is sometimes denied.

There is some positive evidence that the transformation of *Homo erectus* into *Homo sapiens* was both more or less gradual² and occurred in different parts of the world^{3,5}. The latter conclusion has been, and remains, especially common to students of East Asian and Australasian fossils. This evidence cannot simply be ignored with impunity. The phenomenon has been proposed for some other mammals also (refs 6–8 and M. Freudenthal, personal communication).

As discussed in more detail elsewhere^{4,9}, geographically extended phyletic speciation is a single process, not a matter of independent origin in different places in the manner of some polyploid species. The frequency of occurrence of such phyletic speciation remains controversial, but the mechanism itself consists entirely of a combination of generally accepted processes. These are, first, temporal continuity of regionally adaptive features; second, persisting gene flow from region to region; third, a single potential origin for each generally adaptive feature; and fourth, the spread of generally adaptive features by dispersal and natural selection within pre-existing regional populations elsewhere. The possibility of more than one place of partial or even complete origin is plausible for polygenic characters but it is unnecessary because of development or genetic interactions. 'Hitchhiking' and the like can partly homogenize neutral alleles and other DNA geographically. There is no more reproductive isolation at any one time than within normal species today. The above process is called speciation because it creates an entity which is adaptively (and in other respects) as distinct from its ancestor as are the species we see around us from their close contemporaneous relatives. Such differences are how paleontologists recognize allochronic species. If a later population could travel in time back to an earlier one, it might find itself reproductively isolated, but since this is both untestable and without causal effect, it is irrelevant to the pheromone. There is no implication as to how rapidly phyletic speciation may occur.

But one should not ignore the molecular evidence either, which points to differences of frequencies of variants among populations and even some qualitative

differences. These are used to construct distance matrices and phylogenetic trees, not all of which agree¹⁰ even apart from the usual and (at this scale) unjustified^{11–13} assumption of equality of rates of change. However, in the absence of locally temporal calibration (with associated error of estimate), we really have no idea when the branchings occurred. Perhaps they really are different for different genetic elements; under selection, genomes need not disperse even approximately together. They may have occurred before the origin of *H. sapiens*¹⁴, in part during it, afterwards, or all three.

The suggestion^{1,15} of a bottleneck in population size will be more convincing if enough genetic elements unrelated both functionally and chromosomally can be found to provide positive evidence. A bottleneck is now merely a plausible alternative; processes like meiotic drive and individual selection can mimic its results in particular cases, and different genetic elements now suggest different scenarios.

LEIGH M. VAN VALEN

Biology Department (Whitman),
University of Chicago,
915 East 57th Street,
Chicago, Illinois 60637, USA

1. Jones, J.S. & Rouhani, S. *Nature* **319**, 449–450 (1986).
2. Wolpoff, M.H. *Paleobiology* **10**, 389–406 (1985).
3. Wu, R. & Dong, X. in *Paleoanthropology and Paleolithic Archaeology in the People's Republic of China* (eds Wu R. & Olsen, J.W.) 79–89 (Academic, Orlando, 1985).
4. Wolpoff, M.H., Wu, X.Z. & Thorne, A.G. in *The Origins of Modern Humans* (eds Smith, F.H. & Spencer, F.) 411–483 (Liss, New York, 1984).
5. Smith, F.H. *Z. Morph. Anthro.* **75**, 197–222 (1985).
6. Martin, R.A. *Science* **167**, 1504–1506 (1970).
7. Martin, R.A. *Evol. Monogr.* **2**, 1–36 (1979).
8. Heintz, E. *Bull. Soc. géol. Fr.* **16** (7), 411–417 (1974).
9. Van Valen, L.M. *Perspect. Biol. Med.* **9**, 377–383 (1966).
10. Wainscoat, J.S. *et al. Nature* **319**, 491–493 (1986).
11. Gillespie, J. H. *Proc. natn. Acad. Sci. U.S.A.* **81**, 8009–8013 (1984).
12. Van Valen, L.M. *J. molec. Evol.* **3**, 89–101 (1974).
13. Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C. & Cavalli-Sforza, L.L. *J. molec. Evol.* **19**, 255–271 (1983).
14. Cann, R.L., Brown, W.M. & Wilson, A.C. *Genetics* **106**, 479–499 (1984).
15. Brown, W.M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3605–3609 (1980).

The origins of chloroplasts in eukaryotes

SIR—While I agree with most of A.E. Walsby's *News and Views* comment¹ on the significance of the discovery by Burger-Wiersma *et al.*² of a filamentous prochlorophyte, I must take issue with his statement that the pairing or stacking of thylakoids (chlorophyll-bearing membranes) "... may be a quasi-mechanical consequence of the presence of chlorophyll *b* and the absence of phycobilisome structures ..." and "may not therefore be of any phylogenetic significance". This neglects a fundamental difference in the organization of photosynthesis between cyanobacteria (and eukaryotic red algae) on one hand and prochlorophytes and green eukaryotes on the other.

In cyanobacteria the thylakoid membranes are laterally homogenous; both photosystem II (which seems to be specifically associated with the phycobilisomes) and photosystem I are distributed across the whole membrane. In green eukaryotes there is a segregation of function; photosystem II is largely or entirely concentrated in appressed membrane regions, while photosystem I is predominantly located in single membranes. Freeze-fracture electron microscopy shows that *Prochloron* thylakoids are laterally heterogeneous, with a very similar pattern of intramembrane protein particles to that seen in green algae and higher plants^{3,4}. There is no *a priori* reason why either the presence of chlorophyll *b* or the absence of phycobilisomes would lead to this segregation and membrane specialization; the phylogenetic significance lies in the fact that *Prochloron* both has the same chlorophylls as higher plants and organizes its photosynthesis in the same way. It is conceivable that chlorophyll *b* could have arisen more than once, but hard to imagine that each time it would have been accompanied by the same major restructuring of the photosystem organization in the membrane.

Prochloron resembles the green chloroplast in another feature, the dispersed arrangement of its DNA⁵; cyanobacteria almost always have their DNA in a central nucleoid⁶. Burger-Wiersma *et al.*² report that their organism has a central nucleoid like that of cyanobacteria and unlike other prochlorophytes. They do not mention whether or not the thylakoids are stacked, and their published micrographs do not establish this clearly. Further electron microscopy to settle this point and freeze-fracture studies to establish whether or not the membranes are laterally heterogeneous are the next steps towards understanding the position occupied by *Prochloron* in the evolution of photosynthesis.

GUY COX

Electron Microscope Unit,
University of Sydney,
NSW 2006, Australia

1. Walsby, A.E. *Nature* **320**, 212 (1986).
2. Burger-Wiersma, T. *et al. Nature* **320**, 262 (1986).
3. Giddings, T.H., Withers, N. & Staehelin, L.A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 352 (1980).
4. Cox, G.C. & Dwart, D.M. *New Phytol.* **88**, 427 (1981).
5. Coleman, A.W. & Lewin, R.A. *Phycologia* **22**, 209 (1983).
6. Coleman, A.W. *J. Phycol.* **21**, 371 (1985).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*, but those that do should not be highly technical comments on Articles or Letters (where the Matters Arising section remains appropriate). □

Model of reduction

Philip Kitcher

Neurophilosophy: Toward a Unified Science of the Mind–Brain.

By Patricia Smith Churchland. MIT Press: 1986. Pp.546. \$27.50, £27.50.

THE scholarly reading population divides naturally into two types. There are absorbers, people who like to be informed and who delight in the lucid presentation of unfamiliar ideas, facts and theories. Arguers, by contrast, like to fill the margins of their best-loved books with questions and comments (usually prefaced by "Yes but ...") and they prize ingenious defences of controversial theses. Patricia Churchland's new book should appeal to readers of both types.

Neurophilosophy offers an illuminating survey of contemporary neuroscience, and a review of recent developments in the philosophy of science and the philosophy of mind, which lead Churchland to a distinctive philosophical position and to a series of suggestions about how scientists and philosophers together might develop large explanatory theories of brain functioning. A leading theme of the entire book is that, in Quine's famous phrase, "there is no first philosophy". Philosophical reflection about the nature and function of minds must be informed — and may be transformed — by the findings of the sciences. Conversely, philosophical synthesis of the ideas of various disciplines can prove profitable to the working scientist, to the neuroscientist as well as the psychologist, by removing conceptual blinkers and opening up new vistas for research. Churchland's philosophical naturalism may be anathema to those whose livelihood depends on keeping

consider, the curious absorber who can't tell the hypothalamus from the hippocampus. The result is a brilliant piece of exposition that should be read by anyone who is interested in what we currently know of the brain and its workings but who does not have the time to invest in a degree course in neuroscience. In my own experience, attempts to explain to non-professionals the intricacies of neuroanatomy and neurophysiology typically pile detail upon detail, technical term upon technical term, until the bewildered reader is lost in the mossy fibres. Churchland's account proves entertaining instead of soporific, partly because of her lively style, partly through the telling use of illustrations, but chiefly because of her organization of the material. What might have been a catalogue of cell types, neurotransmitters and brain areas becomes an exciting travel guide, in which we are informed both about the history of the discoveries concerned and about the nature (and limitations) of methods for acquiring information about the brain.

In the second part of her book, Churchland turns her attention to the philosophy of science and the philosophy of mind. She begins with an overview of twentieth-century philosophy of science, in which she outlines the principal themes of logical empiricism and chronicles its death (or transfiguration?) under the impact of the ideas of Kuhn and Feyerabend. Her discussion sets the stage for a more intensive examination of the concept of "inter-theoretic reduction". Churchland provides a model of reduction according to which theories are viewed as sets of sentences and reduction consists in the deduction of *analogues* of the statements of one theory (the reduced theory) from the principles of the other (the reducing theory).

The concept of reduction is important to Churchland because she sees the issue of the reducibility of psychology to neuroscience as *the* question in the philosophy of mind. Furthermore, to anticipate a little, finding an appropriate answer to this question and integrating it with our current neuroscientific knowledge might produce, as Churchland contends in the final chapter of her book, both improvements within neuroscience and a transformation of philosophy. The fundamental issue is approached carefully. After a penetrating discussion of varieties of mind–brain dualism — which will, I hope, diminish the fascination that some

neuroscientists find in these hopeless doctrines — Churchland focuses on what she sees as the most sophisticated anti-reductionist position, the thesis that psychological states are *functional* states (that is, states that occupy a particular functional role in the life of an organism) which have extremely heterogeneous neurological realizations. Her critique of this position is integral to her development of her own views about psycho-physical reduction.

The kernel of Churchland's intricate and wide-ranging argument runs as follows. We should not expect neat reductions that provide simple and uniform identifications of the entities and properties discussed by the reduced theory using language from the reducing theory, and that derive "corrected versions" of the principles of the reduced theory which introduce only *minor* corrections. For such neat reductions are not to be found among the philosopher's paradigm cases from physics, let alone in the example, which Churchland takes to be more pertinent, of genetics. Moreover, even if the distinctions drawn by psychology should cut

"How can we make sense of mental activities without employing the ideas of folk psychology and assuming that thought involves the manipulation of symbols?"

across the divisions of neuroscience, this does not signify that psychology will persist as an autonomous science remaining irreducible to neuroscience. Rather, we ought to expect psychology and neuroscience to coevolve, so that their categories come to mesh with one another. Although this may demand the sacrifice of some everyday psychological notions, Churchland contends that *folk psychology* — roughly that body of lore that we bring to the understanding and prediction of the behaviour of others — is in trouble anyway, and, in this connection, she cites the recent work of Stephen Stich and Daniel Dennett. If I interpret her correctly, the argument is clinched by appeal to the unity of science. Functionalists who flaunt the irreducibility of psychology are to be hoist with their own petard. Unity of science demands that the barriers to reduction must be removed, and psychology (or neuroscience) must amend itself to suit.

I am not convinced by this line of reasoning. One source of my resistance lies in the fact that, while she is happy to advocate conceptual reform for psychologists, and though she explicitly considers the possibility that science might lead us to modify our philosophical concepts, that possibility is not taken seriously in the present context. Perhaps Churchland is right in thinking that something like the

"Churchland writes with the authority of an insider; one who has spent nearly a decade not only in polite conversation with neuroscientists but also with her hands in the brain."

philosophy pure, but it harmonizes beautifully with the efforts of other contemporary philosophers who have begun (or continued) a dialogue with one or another of the sciences. Moreover, *Neurophilosophy* provides an admirable illustration of how fruitful the exchange may be.

The first 235 pages present the main achievements of neuroscience past and present. Churchland writes with the authority of an insider, one who has spent nearly a decade not only in polite conversation with neuroscientists but also with her hands in the brain. She also understands clearly the predicament of the out-

question of reducibility of psychology to neuroscience is the fundamental problem in the philosophy of mind. Yet the analogy I draw from the case of genetics is that the philosophical concept of reduction is inadequate to the representation of the relations among many theories (or fields) of science*. To borrow Churchland's own terminology, I think that the concept of reduction may belong to *folk metascience*.

A second concern about her argument derives from a different way of appealing to the unity of science. One of the leading themes of functionalism is the view that there are generalizations about psychological processes — couched in terms of representation and computation — which apply to systems that are built out of very different materials. Assuming that such generalizations cannot be integrated into neuroscience, the functionalist may reasonably protest against their sacrifice on the altar of the unity of science *on the grounds that they already contribute to a unified science by linking together the study of different forms of intelligence both artificial and natural*. Churchland's apparently clinching argument thus seems to presuppose a particular view of how an ideally unified science must be organized, a view on which neuroscience stands as the neighbouring discipline to psychology. Functionalists who believe in unified science should adopt a different picture of its structure, and they may accuse Churchland of extracting more from a methodological principle than it can fairly yield.

These reservations do not detract from the achievement of the second part of *Neurophilosophy*. Whether or not one agrees with her, Churchland has performed a great service for philosophers, psychologists and neuroscientists through her elaboration of the major positions and her clear presentation of arguments. Equally, the long and fascinating chapter that makes up the third and final part of the book, should serve as the starting point for serious interdisciplinary discussions.

We learned in Churchland's review of neuroscience that the various branches of the field are urgently in need of theory. Attention to the philosophy of mind offered the moral that philosophical conceptions, including those that strike us as untutored common sense, may have to be "reconfigured". But what are the prospects for theories in neuroscience? And how can we make sense of mental activities without employing the ideas of folk psychology and assuming that thought involves the manipulation of symbols? Churchland's final chapter outlines some

answers to these questions. She gives a lucid account of some recent research: the tensor network theory of Pellionisz and Llinas (and a related proposal by Paul Churchland), the parallel models of computation developed by Rumelhart and McClelland, and a hypothesis of Crick's about the neurobiology of attention. In each case, there appears to be an exciting symbiosis between naturalistic philosophy and theoretical neuroscience, with the liberation from conceptual chains leading to new prospects for developing theory.

Tensor network theory receives the most extended treatment. The core of the Pellionisz-Llinas proposal is that synaptic connections in the cerebellum can be viewed as an array for performing matrix multiplication. So, to simplify enormously, sensorimotor coordination might be achieved by means of the registering of information on a sensory map in the brain, the "computation" in the cerebellum of an appropriate vector in a motor map, and, as the result of the stimulation of the right region of the motor map, the activation of the proper muscles. The proposal is especially stimulating because the "computation" does not involve manipulation of symbols. Algebra gives way to geometry. Churchland offers a succinct summary:

What is needed is a way to conceive of what nonsentential representing might be, and the tensor network theory provides that much, even if, in the end, it turns out not to be right [p.452].

Functionalists will be quick to protest that managing without symbol manipulation in the case of sensorimotor coordination is

one thing, doing the same in the case of language learning quite another. Neuroscience can cope with the more mundane operations of the brain, but the higher cognitive functions require a more abstract treatment. Taken as a cautionary note, this is perfectly correct, but the point should not be overinterpreted. The work Churchland describes is extremely tantalizing, for neuroscience, for psychology and for philosophy. We do well to recognize exactly how far we have gone, but not to make unwarranted judgments about how far we may go.

Churchland is plainly optimistic, and her optimism may ultimately prove unjustified. But, whatever the outcome of the trends in theoretical neuroscience that she favours, the value of her book is independent of its more speculative claims and its more controversial arguments. While reading it, I was reminded of the comments of numerous friends, who mourn the good old days — days when there were far fewer academic books, when books were read by several generations of graduate students, when the brief bright ideas emerged on journal pages but not in the more enduring medium of hard covers. Such people should be happy with Churchland, for she has given them a book of the old-fashioned sort, one that deserves to be read and pondered by philosophers, psychologists and neuroscientists for a good long time. □

Philip Kitcher is Director of the Minnesota Center for the Philosophy of Science, 315 Ford Hall, 224 Church Street SE, Minneapolis, Minnesota 55455, USA. His most recent book is Vaulting Ambition: Sociobiology and the Quest for Human Nature (MIT Press, 1985).

Revival in optics

O.S. Heavens

Optical Interferometry. By P. Hariharan. Academic: 1986. Pp.303. \$58, £49.50.

OPTICS — until recently perhaps regarded as satisfactorily completed and with little more to be expected — has experienced a resurgence of interest since the arrival of the laser, a tool of almost infinite sharpness compared with the light sources previously available. The laser has not only made many traditional optical techniques more convenient, but has made possible methods such as holography which were hitherto practically impossible.

An account of optical interferometry can thus be expected to contain some new excitements to complement the more traditional material. In this sense, Hariharan's book is no disappointment: holography, holographic interferometry and speckle interferometry — all post-laser subjects — are well covered in a mainly

readable fashion. There are some grumbles. The pedagogue tires of explaining to students why the term "phase of a photon" is completely meaningless, but here it is again. There is, too, an uneven quality to the book in that rather basic algebra is spelled out at length while a knowledge of Jones matrices is assumed. Some carelessness in checking the matching of symbols in text and figures has introduced glitches in reading, and many of the line diagrams could have been made clearer by the use of shading.

One instinctively draws a comparison between this book and the classic work of Steel, *Interferometry*, of which a new edition appeared in 1983 and which is now available in paperback (Cambridge, £12.95, \$24.95). They have much in common, since Steel's updating of the earlier (1967) book was thorough. The better diagrams of Steel's version and freedom from some of the blemishes referred to above, will probably mean that its sales will not suffer. □

O.S. Heavens is Professor of Physics, University of York, York YO1 5DD, UK.

*For a study of the example of genetics that seeks to develop more sophisticated meta-scientific concepts, see my article "1953 and All That. A Tale of Two Sciences" (*Philosophical Review* 93, 335–373; 1984)

How does the brain do it?

Stuart Sutherland

Visual Cognition. Edited by Steven Pinker. MIT Press: 1986. Pp.270. Pbk \$17.50, £17.50.

MANY years ago, I put forward the following problem. A cross and a circle, horizontally aligned, are flashed onto the retina; regardless of the shapes' absolute positions, the subject has to determine which is on the left. As visual tasks go, it could hardly be simpler: the problem is that we have no idea how the brain does it. A scan from left to right (or vice versa) seems improbable. If, on the other hand, the coordinates of each shape are carried forward, how is this done and how is the comparison between the coordinates made in neural terms?

In the most original and important article in *Visual Cognition*, Shimon Ullman points to a series of similar problems. He asks what mechanism could trace out a closed curve to decide that it is closed and to determine which points lie inside and which outside it. The problem is a more precise version of the figure-ground effect, so beloved by the Gestaltists. Although the question can be stated more rigorously than in their day, we are no nearer a solution, at least in terms of brain mechanisms, for there is no difficulty in writing computer programs to carry out this task.

At the other extreme of vision are complicated problems which still defy computers, but which are solved by the brain by means that remain obscure. Perhaps the most difficult is that of visual recognition. Steven Pinker gives a judicious review of the subject, noting the many difficulties that beset current theories. We have of course moved on: nowadays Gestalt theorizing looks hopelessly vague, and the idea that recognition could be achieved by templates or by feature matching, which was being put forward in all seriousness only 20 years ago, now looks extremely naive. Some of the more recent ideas will probably survive, for example that, for the purposes of recognition, the brain builds an abstract and hierarchical structural description of an object. However, the exact nature of such descriptions, the methods by which they are derived from the retinal image, and how or indeed whether the same description is formed of an object regardless of viewpoint remain entirely open questions.

Half of the papers collected by Pinker are devoted to visual imagery and the ways in which people manipulate it, a subject on which there has recently been much research. There are a limited num-

ber of transformations that can be applied to the image of an object — it can, for example, be made to change size, to rotate or to change position. In all such changes, the imaged object passes through all intermediate points. One cannot suddenly replace an image of an upright letter with that of the same letter tilted through 90°. Steven Kosslyn has constructed an ingenious computer model which incorporates the processes required to manipulate imagery in the way people do. He describes his model and the results of submitting it to a series of tests, in which he uses consistent differences in people's ability to perform the different operations to demonstrate that they really are used by the brain.

Such questions as whether the tail of a horse ends above or below its knees can only be answered by using imagery. But if for the purpose of recognition we construct structural descriptions, it is puzzling that the answer cannot be read off directly from them. Perhaps this relationship is only implicit in the description, making it necessary to generate an image from the structural description and derive the answer from that.

The present volume, originally a special issue of the journal *Cognition*, is useful in so far as it delineates the state of the art in a few branches of vision. It makes it clear that we have advanced mainly by discovering and making precise an increasing number of problems, and that theorizing is more rigorous than ever before. But although the secret of how the brain subserves vision has for some time appeared to be just around the corner, it still remains elusive. How, for instance, does my brain calculate that my typewriter lies to the left of my gin bottle? □

Stuart Sutherland is Director of the Centre for Research on Perception and Cognition, University of Sussex, Brighton BN1 9QG, UK.

Morphological eye

Malcolm Irving

Skeletal Muscle. Handbook of Microscopic Anatomy, Vol.2, Part 6. By Henning Schmalbruch. Springer-Verlag: 1985. Pp. 440. DM 580.

THE PREVIOUS edition of this handbook was published in 1956, a time of revolution in muscle research. The idea that contraction involved the relative sliding of two sets of filaments was new, and the first clear electron micrographs of the side-by-side array of interdigitating filaments were still to be published. In the 30 years since then electron microscopes seem to have been focused on every structural feature in muscle, so that we now have a fairly complete picture of its ultrastructure, at

least down to a resolution of about 10nm. Parallel progress with other techniques has in many cases helped to reveal the functional significance of the structural specializations.

This new edition is therefore long overdue and, in attempting to review progress since the last one, Schmalbruch was confronted with an embarrassment of riches. He has reduced his subject to manageable proportions by concentrating on one approach to the study of muscle, which he describes as "muscle seen through the eyes of a morphologist". This allows him to treat a wide range of biologically interesting phenomena from a unified viewpoint. The morphological aspects of neuromuscular transmission, contractile proteins and the mechanism of force generation, internal membranes and the regulation of contraction, metabolism and fibre types, development and regeneration are all dealt with in some depth. A wide range of other features, for example connective tissue, vascularization and the organization of motor units, are described more briefly. The emphasis is on mammalian muscle (including a section on non-skeletal striated types, but excluding cardiac muscle), though results from studies on other animals are discussed where they serve as useful experimental models. Pathological states are mentioned in passing but not treated systematically.

Schmalbruch has painstakingly organized and cross-referenced his material, which is also beautifully illustrated by over 200 micrographs and backed up by about 2,000 references. He does not hesitate to pepper the text with quantitative detail where it is available (thus we learn that a 10cm cell from a human biceps contains 3,000 nuclei), or to resort to descriptions of "amorphous densities" or "ill-defined periodic structures" where it is not. Technical issues, such as possible problems in the use of a particular fixative, are discussed where they might affect the interpretation of the literature. The rather empirical "morphologist's eye" approach adopted for most of the book often leaves the reader to draw his or her own conclusions from the data; sometimes, though, as in the section on myogenesis, Schmalbruch allows himself to discuss alternative hypotheses, giving freer rein to the morphologist's brain.

This book is more than an atlas of muscle structure, but less than a complete account of the structural basis of muscle function, and it will probably be of more use to the research worker than to the undergraduate. It should remain a useful reference volume for a long period — if not quite until the publication of the next edition, which at the present rate should appear sometime after the year 2010. □

Malcolm Irving is in the Department of Biophysics, King's College London (KQC), 26-29 Drury Lane, London WC2B 5RL, UK.

Neolithic battles

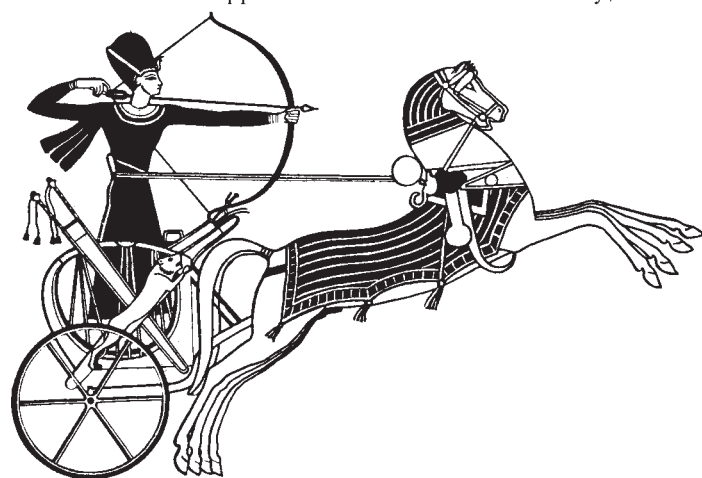
William H. McNeill

The Origins of War: From the Stone Age to Alexander the Great. By Arther Ferrill. *Thames & Hudson: 1985. Pp.240. Hbk £12.50, \$19.95; pbk £5.95, \$10.95.*

ARTHER Ferrill is a professor of ancient history at the University of Washington, Seattle, and in this book he advances two general propositions that are likely to surprise most readers, as they did me.

First, he describes a Neolithic "burst of organized warfare" in the Near East which, he argues, was "as important for the area as Alexander's conquest of Persia in the fourth century BC or the march of Islam in the seventh century AD" (pp. 27–28). New weaponry was part of this change: the bow, the sling, the dagger and the mace all start to appear in the archaeo-

logical record between 12000 and 8000 BC. Regular military formations (column and line) made the use of such weapons more effective, and elaborate fortifications had to be built to protect against organized attack — all before settled agriculture had become widespread. Indeed, Ferrill suggests that fixed residence, dictated by the need for protective walls, may have led to the development of fixed agriculture (p.29), without, however, pressing the point as to which came first.



Ramses II in his war chariot at the Battle of Kadesh in 1285 BC: his lighter chariots and strategy gave him a tactical advantage against the Hittites. Ramses' strategy and tactics showed a degree of military sophistication that was not seen in many later historical periods.

Ferrill's second proposition is that classical Greek warfare was backward and technically deficient compared with the military techniques and command system of the Persian empire. That empire, heir to Bronze Age empires and to the even more impressive Assyrian techniques of waging war, added a navy to combined-arms warfare on land, and could in addition support both branches with an adequate logistical system. By comparison, the Greek style of hurly-burly duelling between aristocratic horsemen (who Ferrill says, dismounted to fight in the eighth century: hence Homer's chariot tactics!) and the unsupported charge of

superiority in the fifth and fourth centuries requires him to explain away Greek victories on the plain of Marathon and at Plataea. His schematized accounts of the battles leave aside all the innumerable contested issues as to what actually happened. If Persian combined arms were so superior, why did they not prevail on these occasions? Ferrill's brief battle narratives simply do not show "that the Persians were defeated by their own mistakes", although he declares that this view "must surely be correct" (p.122).

Throughout, the author uses the style of an essayist, and simply asserts what must often be guesswork. He reaches for novelty of interpretation, and perhaps overreaches himself. But as long as one understands what treacherous ground he traverses, and how controversial most of his assertions are, the book makes good reading; and I, for one, was persuaded that the Neolithic "military revolution" in the Near East was as real and important as Ferrill says it was. □

William H. McNeill is a Professor in the Department of History, University of Chicago, 5801 S. Ellis Avenue, Chicago, Illinois 60637, USA.

Good and the bad

John Crothers

The Encyclopaedia of Reptiles and Amphibians. Edited by Tim Halliday and Kraig Adler. **The Encyclopaedia of Insects.** Edited by Christopher O'Toole. *George Allen & Unwin/Facts on File: 1986. Both: Pp.152. £15, \$22.50.*

THESE are the fifth and sixth volumes in the deservedly successful Unwin Animal Library. Previous titles have dealt with mammals (two volumes), birds and under-water life. The features uniting all these works are a systematic format, authoritative, concise text, and superb illustrations; all are beautifully produced and are a pleasure to read.

The Encyclopaedia of Reptiles and Amphibians must be the best book on herpetology currently available to the general reader — and there can be few professional zoologists working in related fields who will not welcome it onto their shelves. The extraordinary collection of colour photographs is augmented with quite excellent paintings. But, unlike many beautiful books, the text would stand on its own. It is arranged ordinarily, with a small box at the beginning of each section describing the main features of the order (or sub-order) concerned — the number of species, their distribution, habitats, size, colour, reproduction and longevity. There are also separate sections on parthenogenesis, shedding of lizard tails and so on.

I was initially surprised to find no mention — in the systematic part — of the influence of temperature on sex determination in crocodilians, but the topic is discussed earlier in the general introduction to reptiles. Unlike *The Encyclopaedia of Mammals*, this volume does not include an appendix listing species, genera and families, and giving both the scientific and colloquial names. But that is just about my only criticism.

By comparison, the volume on insects is disappointing. The title is misleading, as the work seeks to cover most of the terrestrial arthropods, and for some reason it has been decided that all these invertebrates only justify 143 pages of text (the same as amphibians and reptiles), against 895 for the mammals. The 9,000 species of reptiles and amphibians have an entire book to themselves but the 11,500 species of millipedes, centipedes and their allies are covered in just four pages of this one. Even the 300,000 known species of Coleoptera (beetles) are given only twelve. The result, inevitably, is a very much more superficial account. □

John Crothers is Warden of The Leonard Wills Field Centre, Nettlecombe Court, Williton, Taunton, Somerset TA4 4HT, UK.

Kelvin on an old, celebrated hypothesis

from Edward Harrison

Lord Kelvin in 1901 tested an "old and celebrated hypothesis" that if we could see far enough into space the whole sky would be occupied with stellar disks all of perhaps the same brightness as the Sun. Kelvin was the first to solve quantitatively and correctly the riddle of a dark night sky, a riddle that had been previously solved qualitatively by Edgar Allan Poe, and is now known as Olbers' paradox.

In a paper¹ published in 1901, Lord Kelvin discussed various astronomical topics, such as the size and mass of the Galaxy, the velocities of stars, and the dark night-sky riddle. Kelvin's treatment of the riddle of darkness at night is of considerable scientific and historical interest; for the first time in the context of a static universe he solved the riddle quantitatively and correctly. This paper was reproduced in 1904 as the revised Lecture XVI in the *Baltimore Lectures*². The omission of this paper from Kelvin's *Mathematical and Physical Papers*³ and from the complete bibliography of his works in Silvanus Thompson's *Life of William Thomson*⁴ may account for its neglect in numerous subsequent discussions of the so-called Olbers' paradox⁵⁻⁷.

The riddle of a dark starlit sky can be traced back to Thomas Digges in 1576 and Johannes Kepler in 1610⁸. The modern expression 'Olbers' paradox' gained popularity in ignorance of these contributions, and of those made by Otto von Guericke⁹ in 1672, Edmund Halley¹⁰ in 1721, Jean-Philippe Loys de Cheseaux¹¹ in 1744, and many others, all of which preceded the work by Wilhelm Olbers in 1823¹².

The riddle is easily stated in the form proposed by Olbers. Plausibly, space is unbounded and stars are scattered throughout space; a line of sight extended in any direction ultimately intercepts the surface of a star. Every point of the sky must blaze with starlight with no dark gaps between the stars. If most stars are like the Sun, the sky at every point must be as bright as the Sun's disk. The sky subtends a solid angle 180,000 times that of the Sun, so the starlit sky should be 180,000 times brighter than the Sun. But the sky at night is dark—hence the riddle.

Kelvin's analysis

The following parallels the analysis given by Kelvin. We assume for computational convenience that all stars are Sun-like, of radius a , and uniformly distributed. Let n be their number per unit volume. Then the number in a shell of radius q and thickness dq is $4\pi nq^2 dq$. This number multiplied by $a^2/4q^2$ represents the fraction of the whole sky covered by stars viewed by an observer at the centre. To include geometric overlap of stellar disks, neglected by Kelvin, we

multiply by $\exp(-q/\lambda)$, where $\lambda = (\pi na^2)^{-1}$ is the mean free path of a light ray. Hence the fraction of the sky covered by stars in the shell is $\exp(-q/\lambda)dq/\lambda$. By integrating from $q=0$ to $q=r$ we obtain

$$\alpha = 1 - e^{-r/\lambda} \quad (1)$$

where α denotes the fraction of the sky covered by stellar disks out to radius r . In a distribution of stars of infinite extent the entire sky is covered ($\alpha=1$) and the average distance seen from any position is the background limit λ . We arrive at Kelvin's result (his equation (10)) when the radius r of the distribution of stars is small compared with the background limit, so that

$$\alpha \approx \frac{r}{\lambda} = \frac{3N}{4} \left(\frac{a}{r} \right)^2 \quad (2)$$

where $N = 4\pi r^3/3$ is the total number of stars. This treatment by Kelvin elucidates that previously given by Halley, Cheseaux and Olbers.

Let each star have luminosity L . The contribution to the radiation density u at the centre from the stars in a shell is $(nL/c)e^{-q/\lambda}dq$, where c is the speed of light. By integrating as before we find

$$u = u^*(1 - e^{-r/\lambda}) \quad (3)$$

where $u^* = L/\pi a^2 c$ is the radiation density at the surface of a star¹³. Hence $u = u^*$ in any distribution of stars extending much beyond the background limit.

From equations (1) and (3) we obtain

$$\alpha = u/u^* \quad (4)$$

thus demonstrating the truth of Kelvin's remark that " α is the ratio of the apparent brightness of our star-lit sky to the brightness of our Sun's disc." In many recent discussions of the riddle this significant relation has received scant recognition.

Solutions

With his model of the Galaxy, consisting of a sphere of radius 10^3 pc (1 pc = 3.26 light-yr) containing 10^9 Sun-like stars, Kelvin found $\alpha = 4 \times 10^{-13}$, and said, "This exceedingly small ratio will help us to test an old and celebrated hypothesis that if we could see far enough into space the whole sky would be seen occupied with discs of stars all of perhaps the same

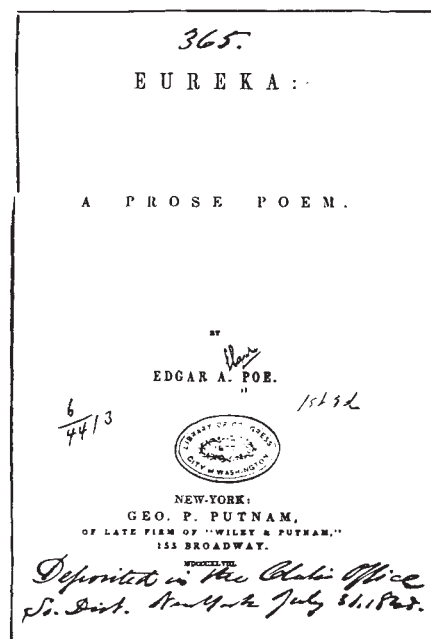


Fig. 1 The title page of Edgar Allan Poe's *Eureka*, published in June 1848, dedicated to Alexander von Humboldt, and based on a two-hour lecture, "The cosmogony of the universe", delivered at the Society Library, New York, in February 1848 while the Rev. Dr John Pringle Nichol, regius professor of astronomy at the University of Glasgow, was also lecturing in New York.

brightness as our own Sun."

We require, said Kelvin, a "great sphere" of stars, of radius λ ($\approx 10^{14}$ pc), to cover the sky with stellar disks. But starlight travelling from the outlying stars of such a great sphere takes 3×10^{14} yr to reach us. Using the theory that stars derive their energy by gravitational contraction, he said, we know that 10^8 yr is an upper limit for a stellar luminous lifetime. "Hence, if all the stars through our vast sphere commenced shining at the same time... at no one instant would light be reaching the Earth from more than an excessively small proportion of all the stars." In this remarkable argument Kelvin showed that the transit time of light from the background limit greatly exceeds the luminous lifetime of stars. If all stars began to shine at the same instant, we see light only from stars in the foreground, not from those in the distant background.

Relativity theory had yet to stress the

universality of $E = mc^2$, which provides a true maximum luminous lifetime $M_\odot c^2/L_\odot$ of $\sim 10^{13}$ yr for the Sun. This extreme value is still less than Kelvin's limit of 3×10^{14} yr and could not have altered his conclusion. It was not evident in Kelvin's day that the universe contains insufficient energy to create a bright starlit sky¹⁴. We now know that the conversion of all mass in the universe into thermal radiation would attain at most a temperature of 20 K, much less than the Sun's surface temperature of 6,000 K.

We may re-write equation (3) in the form:

$$u = u^*(1 - e^{-t/\tau}) \quad (5)$$

where $t = r/c$ is the transit time of light from stars at distance r , and $\tau = \lambda/c$ is the transit time from the background limit. When α is small,

$$\alpha = u/u^* = t/\tau \quad (6)$$

thus confirming Kelvin's conclusion that the sky is dark because the transit time τ from the background limit greatly exceeds the maximum value of t equal to the stellar luminous lifetime. Equation (5) can be derived thermodynamically by considering the radiation density in a cavity containing luminous sources; t is then the timespan during which the sources emit radiation and τ the characteristic time required to fill the cavity with radiation in equilibrium with the sources¹⁴.

Edgar Allan Poe was perhaps the first to give a correct qualitative solution to the riddle of darkness. In 1845 he wrote¹⁵,

Look down into the abysmal distances!—attempt to force the gaze down the multitudinous vistas of the stars, as we sweep slowly through them thus—and thus—and thus! Even the spiritual vision, is it not at all points arrested by the continuous golden walls of the universe?—the walls of the myriads of the shining bodies that mere number has appeared to blend into unity?

Three years later, in his masterpiece *Eureka*¹⁵, he formulated his most radical cosmological ideas. Concerning the riddle of darkness, he said,

Were the succession of stars endless, then the background of the sky would present us an uniform luminosity, like that displayed by the Galaxy—since there could be absolutely no point, in all that background, at which would not exist a star. The only mode, therefore, in which, under such a state of affairs, we could comprehend the voids which our telescopes find in innumerable directions, would be by supposing the distance of the invisible background so immense that no ray from it has yet been able to reach us at all.

A few lines he wrote, "I myself feel impelled to the fancy—without daring to call it more—that there *does* exist a *limitless* succession of Universes, more or less similar to that of which we have cognizance." By "Universes" he meant the

BALTIMORE LECTURES

ON

MOLECULAR DYNAMICS

AND

THE WAVE THEORY OF LIGHT

FOUNDED ON MR A. S. HATHAWAY'S STENOGRAPHIC REPORT OF
TWENTY LECTURES DELIVERED IN JOHNS HOPKINS
UNIVERSITY, BALTIMORE, IN OCTOBER 1884;
FOLLOWED BY TWELVE APPENDICES ON ALLIED SUBJECTS

BY

LORD KELVIN, O.M., G.C.V.O., P.C., F.R.S., &c.

PRESIDENT OF THE ROYAL SOCIETY OF EDINBURGH,
FELLOW OF ST JOHN'S COLLEGE, CAMBRIDGE,
AND FRIESTER PROFESSOR OF NATURAL PHILOSOPHY IN THE UNIVERSITY OF GLASGOW.

LONDON:

C. J. CLAY AND SONS,
CAMBRIDGE UNIVERSITY PRESS WAREHOUSE,
AVE MARIA LANE,
GLASGOW: 26, WELLINGTON STREET.

BALTIMORE:

PUBLICATION AGENCY OF THE JOHNS HOPKINS UNIVERSITY.

1904

Fig. 2 Title page of the *Baltimore Lectures*, published in 1904. Kelvin's solution of the riddle of darkness is given in Lecture XVI, pages 260–278. The solution does not appear in the original lecture notes of 1884, and was first published in 1901 in the *Philosophical Magazine* in a paper entitled "On ether and gravitational matter through infinite space".

Galaxy and other galaxies. The finite speed of light and the age of the starlit universe had come together in Poe's fertile mind, shedding new light on an old hypothesis.

Comments

Despite the widespread interest in Olbers' paradox in recent years, little attention has been paid to the pertinent comments in the nineteenth century by William Herschel, Olbers, Nichol, Poe, Thomas Dick, Friedrich Struve, Alexander von Humboldt, John Herschel, Richard Proctor, Hugo von Seeliger, Ellard Gore, Lord Kelvin and others^{6,9}.

John Pringle Nichol, who introduced the word 'evolution' into astronomy and was regius professor of astronomy at Glasgow University from 1836 to 1859, greatly inspired both Kelvin and Poe. Kelvin acknowledged his debt when recalling his student days^{17,18}. Much of Poe's knowledge of contemporary developments in astronomy came from Nichol's influential *Views of the Architecture of the Heavens*^{19,20}, to which he made reference in *Eureka*. Nichol's astronomy lectures stressed the importance of the transit time of light.

Kelvin pointed out that the supposition of uniform density is arbitrary, and "we ought in the greater sphere to assume the density much smaller than in the smaller sphere." Allowing for the reduced density of stars in the universe because of the wide separation between galaxies, the background limit of Poe's golden walls recedes to 10^{23} light-yr. With a typical stellar

luminous lifetime of 10^{10} yr we see that the fractional coverage $\alpha \approx 10^{-13}$ departs little from Kelvin's value.

According to Kelvin the sky at night is dark because the radius of the starlit *visible universe*, equal to the time stars have been luminous multiplied by the speed of light, is small compared with the background limit, where the background limit is the distance to which stars must extend in order to cover the entire sky. Starlight emitted beyond the visible universe has yet to reach us. Kelvin thus solved the riddle in its original cosmological framework: a static universe of uniformly scattered, Sun-like stars. All variants of this primitive standard model seeking to solve the riddle by appeal to absorption, hierarchy, expansion or fatigued light merely achieve a state of greater darkness in a universe already dark.

Bondi⁵ resurrected the riddle within the context of the steady-state theory. Continually created stars shine unceasingly, according to this theory, and expansion maintains starlight at a constant level. Expansion is necessary (but not sufficient) for darkness, and Kelvin's solution fails in the steady-state model because of its non-conservation of energy.

More than once Kelvin said, "paradoxes have no place in science"²¹. He took the rationalist view that paradoxes lie in ourselves and not the external world. It is historically ironic that he was the first to solve quantitatively with utmost lucidity a riddle that later, when his work lay forgotten, became known as Olbers' paradox. □

Edward Harrison is at the Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01003, USA.

1. Thomson, W. *Phil. Mag.* Ser. 6, 2, 161–177 (1901).
2. Thomson, W. *Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light* (Cambridge University Press, Clay & Sons, London, 1904).
3. Thomson, W. *Mathematical and Physical Papers* (Cambridge University Press, 1882–1911).
4. Thompson, S. P. *The Life of William Thomson, Baron Kelvin of Largs* (Macmillan, London, 1910).
5. Bondi, H. *Cosmology* (Cambridge University Press, 1952).
6. Jaki, S. L. *The Paradox of Olbers' Paradox* (Herder, New York, 1969).
7. Harrison, E. R. *Cosmology: The Science of the Universe* (Cambridge University Press, New York, 1981).
8. Harrison, E. R. *Science* 226, 941–945 (1984).
9. Harrison, E. R. *Darkness at Night: The History of a Riddle* (Harvard University Press, in the press).
10. Halley, E. *Phil. Trans. R. Soc.* 31, 22–24; 24–26 (1720–21).
11. Loys de Chéseaux, J. P. *Traité de la Comète*, 223–229 (M. M. Bousquet, Lausanne, 1744).
12. Olbers, H. W. in *Astronomisches Jahrbuch für das Jahr 1826* (ed. Bode, J. E.) 110–121 (C. F. E. Späthen, Berlin, 1823); *Edin. New Philos. J.* 1, 141–150 (1826).
13. Harrison, E. R. *Physics Today* 27, 69–74 (1974).
14. Harrison, E. R. *Nature* 204, 271–272 (1964).
15. Poe, E. A. in *The Science Fiction of Edgar Allan Poe* (ed. Beaver, H.) 171–174 (Penguin, Harmondsworth, 1976).
16. Poe, E. A. *Eureka: A Prose Poem* (Putnam, New York, 1848); reproduced in *The Science Fiction of Edgar Allan Poe* (ed. Beaver, H.) 205–309 (Penguin, Harmondsworth, 1976).
17. Thomson, W. *Nature* 68, 623–624 (1903).
18. Thompson, S. P. *The Life of William Thomson, Baron Kelvin of Largs*, Ch. 1 (Macmillan, London, 1910).
19. Conner, F. W. in *All These to Teach* (eds Bryan, R. A. et al. (University of Florida Press, Gainesville, 1965).
20. Nichol, J. P. *Views of the Architecture of the Heavens* (Dayton & Newton, New York, 1842).
21. Thompson, S. P. *The Life of William Thomson, Baron Kelvin of Largs*, 1125 (Macmillan, London, 1910).

The long and the short of long-term memory—a molecular framework

Philip Goelet, Vincent F. Castellucci, Samuel Schacher & Eric R. Kandel

Howard Hughes Medical Institute, Center for Neurobiology and Behavior, Columbia University College of Physicians and Surgeons, and New York State Psychiatric Institute, New York, New York 10032, USA

A single learning event initiates several memory processes with different time courses of retention. While short term memory involves covalent modification of pre-existing proteins, the finding that long-term memory requires the expression, during learning, of additional genes, makes it possible to analyse in molecular terms the induction and retention of long-term memory.

ANIMALS acquire new information about the world through mechanisms of learning, and retain that information through mechanisms of memory. Clinical and experimental observations indicate that memory is not a unitary process but is thought to have at least two forms, each of which subserves a family of time courses: a short-term form that can last seconds, minutes and hours, and a long-term form lasting days, weeks or years¹⁻⁵. To understand the mechanisms of the major forms of memory, it will be necessary to describe, in cellular and molecular terms, the inductive mechanisms used in the acquisition of information for each form of memory and the various molecular time-keeping steps involved in retaining the resulting memories.

Studies in certain invertebrates (*Aplysia*, *Hermisenda* and *Drosophila*) have shown that the acquisition and retention of information for short-term memory lasting minutes to hours does not require the synthesis of new proteins. Rather, it depends on second messenger-mediated covalent modifications of previously synthesized proteins that modulate the properties of nerve cells and their synaptic connections⁶⁻¹³. Acquisition involves the activation, by neurotransmitters, of receptor-linked enzymes responsible for the synthesis of intracellular messages. These second messengers activate protein kinases that phosphorylate substrate proteins required for the plastic neuronal modifications (Fig. 1, pathway 1). Retention of these plastic changes depends on the duration of the covalent modification of the substrate proteins and the maintained activity of the enzymes responsible for second-messenger synthesis.

Covalent modifications of the sort used in short-term memory have been proposed to become self-reinforcing for certain instances of long-term memory¹⁴⁻¹⁸. However, a variety of studies now indicate that during the acquisition of information whose memory lasts more than 1 day, specific nerve cells need to express genes that are not needed for short-term memory^{5,19-21}. These studies lead us to outline a specific molecular framework for the various time courses of long-term memory. We think it likely that the same modulatory neurotransmitters that act on the cell surface to induce cytoplasmic messengers during learning are not only important for short-term memory, but also lead, through either the same or additional second-messenger systems, to the activation of at least three other overlapping memory processes, each with its own time course of retention (Fig. 1, pathways 2, 3 and 4): (i) A self-reinforcing covalent modification of proteins, encoding for memory lasting many hours. (ii) A transient activation of effector genes for memory lasting more than 1 day, where the time course of memory retention is simply determined by the half-lives of the induced messenger RNAs and proteins. (iii) The induction of a gene sequence for memory lasting weeks or months, in which early regulatory genes lead to maintained alterations in the expression of late, effector genes.

Here, the molecular time-keeping step for memory retention is in the maintenance of an altered mRNA transcription or processing state, much as it is in cellular differentiation.

Time window

Whereas short-term memory does not require the synthesis of new proteins⁶, behavioural studies of learning in vertebrates suggest that memory lasting days or weeks can be disrupted by the inhibition of protein synthesis^{5,19,22,23}. Particularly important for the arguments we develop below is the finding that long-term memory is most sensitive to disruption during and immediately after training. Vertebrates experience no deficit in long-term memory if exposure to the protein synthesis inhibitor is delayed by as little as 1 h after training^{5,19,23}.

This time window for protein synthesis has been found in both lower vertebrates and mammals, and for learning processes that range in complexity from non-associative forms, such as sensitization, to associative forms, such as classical and instrumental conditioning. Although comparable data are lacking in humans, there is the interesting and analogous clinical observation that long-term memory for an experience is particularly susceptible to disruption by convulsions or trauma during and briefly after the experience. Because these findings were based on alterations in behaviour, the studies failed to answer the question of whether this sensitivity to disruption by protein synthesis inhibitors is a systems property of motivation, motor performance or some other complex brain system, or whether it actually reflects a fundamental property of long-term information storage in specific nerve cells of the circuit responsible for the modified behaviour.

This question has now been answered in invertebrates where the neuronal circuitry underlying behaviour and learning is better understood. Studies on long-term sensitization of the gill-withdrawal reflex in *Aplysia* demonstrate that the elementary synaptic connection between sensory and motor neurones, which is enhanced during short-term sensitization, is also enhanced during long-term sensitization²⁴. Examination of this connection in dissociated cell culture has shown that a single application of a facilitating transmitter enhances synaptic transmission for minutes, whereas four or more repeated applications lead to long-term facilitation lasting for over 24 h (ref. 20). In contrast to short-term facilitation at this synapse, the long-term facilitation associated with the memory of long-term sensitization is blocked by both translational and transcriptional inhibitors. As in vertebrates, transient blockage of protein synthesis in *Aplysia* during a 2-h training period is sufficient to block the retention of long-term facilitation assayed 1 day later (refs 20, 21, and V.F.C. *et al.*, in preparation). Similarly, inhibition of both protein and mRNA synthesis after training has little

self-reinforcing mechanism to give a time course intermediate between short- and long-term memory (Fig. 1, pathway 2).

Induced proteins

We believe the sensitivity of long-term memory (lasting more than 1 day) to translational and transcriptional inhibitors during learning is best accounted for by inductive mechanisms, especially because the second messengers induced during learning (and responsible for the shorter forms of memory) are known to be potent gene activators in other systems^{25,26} (Fig. 1, pathways 3 and 4).

Particularly relevant for the discussion of learning are the peptide and biogenic amine hormones that activate receptors linked to enzymes involved in the synthesis of intracellular messengers. The detailed mechanisms by which these second messengers then regulate gene expression are unknown, but it is believed that they modify *trans*-acting regulators of transcription by either leading to their phosphorylation or binding to them directly, in a manner analogous to the binding of CAP protein by cyclic AMP in *Escherichia coli*^{27,28}. The modified transcription regulators could then act on enhancer DNA sequences of structural genes for long-term memory, much as do the DNA-binding proteins that mediate the transcriptional response to steroid hormones or heat shock^{29,30} (Fig. 1, pathway 3). With certain hormones, gene induction is relatively transient and parallels the period during which the hormone is present³¹. The duration of the physiological response is then carried by the stability of the mRNA and the half-life of the protein products. In the case of neurones, where the average half-life of protein is several days, the acquisition process for memory lasting days need, therefore, involve only transient activation of effector genes since the retention of memory could be accounted for by the turnover rate of the induced mRNA and protein products (Fig. 1, pathway 3). These considerations lead us to suggest that it will be important to look for alterations in gene expression, not only during the maintenance phase of memory but, more importantly, during the acquisition phase of memory, that is, during the learning process itself.

One objection to the possibility of genomic regulation of long-term memory that is sometimes raised relates to the question: how can a genomic change lead to long-term alterations in some synapses of a neurone and not others? We suggest that synapse-specific alterations represent a special case of a larger issue frequently encountered in cell biology—the targeting of specific proteins to one rather than another site in the cell^{32–34}. This problem is shared in many other contexts of cell biology. The transient modification at a given synapse produced by short-term memory could allow protein targeting to, recognition by, and stabilization at, that synapse.

Nuclear signals

Memories lasting many days to years clearly persist longer than the half-life of most proteins. In these cases, it is attractive to think of mechanisms for retention that are independent of protein and mRNA turnover, but which depend on persistent, rather than transient changes in transcriptional state, changes that are maintained by cooperative or self-reinforcing mechanisms^{35,36} (Fig. 1, pathway 4). These mechanisms are used in cellular differentiation. Indeed, the often-drawn parallel between long-term memory and cellular differentiation^{37–39} has been strengthened recently by morphological studies of long-term sensitization of the gill-withdrawal reflex of *Aplysia*⁴⁰ and of long-term potentiation in the hippocampus⁴¹ which indicate that long-term memory is associated with growth of synaptic contacts.

Furthermore, the time window so characteristic of memory acquisition is also found in developmental processes in which 'early' genes act as regulators for the expression of 'late' effector genes. For example, the induction of growth and moulting in insects by ecdysone has been shown to alter the expression of

early genes whose induction regulates the expression of late genes. If ecdysone is given together with a protein synthesis inhibitor, the early genes are induced but the late ones are not; if the inhibitor is given several hours later, the late genes are induced normally⁴². A similar class of early genes has been studied more extensively in adenovirus where the nuclear oncogene *E1A*, an early gene, also regulates the expression of the late adenoviral genes⁴³. We suggest that protein synthesis inhibitors not only block the induction of effector proteins for memories lasting a few days, but also block the induction of nuclear signals used by early regulatory genes to activate long-lasting changes in the transcription or processing of the mRNA's of late effector genes. We believe, therefore, that much progress could be made by identifying possible nuclear signals used by early genes to communicate with late effector genes.

Particularly attractive candidates for nuclear signals in long-term memory acquisition are the products of 'competence' or 'immediate early genes'^{44,45}. These proteins are rapidly but transiently induced by hormones and growth factors whose prolonged presentation leads to cellular differentiation and growth. Although these early genes have yet to be shown to regulate the expression of specific effector genes in differentiation, the best-known members of this class, the proto-oncogenes *c-fos* and *c-myc* (which resemble the oncogene *E1A*), have several characteristics which may be relevant to nuclear signalling: (1) the genes respond to the specific cytoplasmic messengers shown to be involved in short-term memory, that is, cAMP, Ca²⁺ and diacylglycerol^{46–48}. (2) The products of these genes are nuclear proteins⁴⁹ that are thought to affect transcription or mRNA processing⁵⁰. (3) The levels of the proteins and mRNA products, when induced, are only transiently increased because of their rapid turnover by degradative processes^{51,52}. (4) The oncogenic forms of these genes are capable of transforming cellular state⁴⁹.

We are not suggesting that *c-fos* and *c-myc* are the nuclear signals responsible for the establishment of memories lasting weeks; but we believe they represent prominent members of a large class of regulatory genes important for development (some of which may be induced during learning), whose protein products may have regulatory functions in the establishment of longer-lasting changes in cellular state. If induced nuclear regulatory proteins important for development prove also to be important for memory acquisition, it would have significance that would extend beyond the biology of learning. It would suggest the existence of common nuclear signalling systems, in the form of mRNA transcription and processing regulators, for long-lived cellular responses much as there are common cytoplasmic signalling systems for transient cellular responses.

Molecular reinforcement

We have proposed that the same extracellular signals that initiate short-term memory (Fig. 1, pathway 1) can initiate, through common intracellular messengers, additional acquisition steps for memory processes with more persistent retention mechanisms (Fig. 1, pathways 2–4). However, not every learning event gives rise to long-term memory. Often a single training trial produces only short-term memory. To produce long-term memory, repeated trials (repeated behavioural reinforcement) are required.

How is information for long-term memory distinguished from that for short-term memory? What are the molecular representations of the behavioural reinforcements necessary for initiating the various long-term memory processes (pathways 3 and 4 of Fig. 1)? To explain the specific activation of these additional acquisition steps, it will be necessary to determine whether the steps require quantitative changes in the levels or durations of the second messengers used in short-term memory, or whether qualitative changes are needed, such as the recruitment of other signalling systems, either cytoplasmic or nuclear. It will therefore be necessary not only to identify and characterize the effector

and regulatory genes whose transcription is activated during the acquisition of long-term memory, but also to delineate necessary molecular reinforcing events. However, the ability to pose some of the critical problems of long-term memory in molecular terms should allow an exploration of its various time courses in a variety of invertebrate and vertebrate learning systems, and thereby make it possible to explore long-term memory at its most fundamental level.

We thank Richard Axel and James H. Schwartz for their useful criticisms, Tom Abrams, Tom Curran, James Darnell, Thomas Jessell, James Morgan, Jeff Morgan and Charles Stevens for their comments on early drafts of this manuscript, Harriet Ayers and Andrew Krawetz for typing the manuscript, Louise Katz and Kathrin Hilten for help with the illustration, and the Helen Hay Whitney Foundation for their support (to P.G.).

1. Ebbinghaus, H. *Memory: A Contribution to Experimental Psychology* (Dover, New York, 1963).
2. James, W. *The Principles of Psychology* (Holt, New York, 1890).
3. McDougall, W. *Mind* **10**, 385-394 (1901).
4. Klatzky, R. L. *Human Memory: Structures and Processes* 2nd edn (Freeman, San Francisco, 1980).
5. Davis, H. P. & Squire, L. R. *Psychol. Bull.* **96**, 518-559 (1984).
6. Schwartz, J. H., Castellucci, V. F. & Kandel, E. R. *J. Neurophysiol.* **34**, 939-963 (1971).
7. Kandel, E. R. & Schwartz, J. H. *Science* **218**, 433-443 (1982).
8. Alkon, D. L. *Science* **226**, 1037-1045 (1984).
9. Farley, J. & Auerbach, S. *Nature* **319**, 220-223 (1986).
10. Livingstone, M. S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5992-5996 (1985).
11. Quinn, W. G. in *Biology of Learning* (eds Marler, P. & Terrace, H.) 197-246 (Dahlem Konferenzen, Berlin, 1984).
12. Dudai, Y. *Adv. Cyclic Nucleotide Protein Phosphoryl. Res.* (in the press).
13. Byrne, J. H. *Trends Neurosci.* **8**, 478-482 (1986).
14. Crick, F. *Nature* **312**, 101-102 (1984).
15. Lynch, G. & Baudry, A. *Science* **224**, 1057-1063 (1984).
16. Saitoh, T. & Schwartz, J. H. *J. Cell Biol.* **100**, 835-842 (1985).
17. Lisman, J. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3055-3057 (1985).
18. Miller, S. G. & Kennedy, M. B. *Cell* (in the press).
19. Agranoff, B. W. *The Chemistry of Mood Motivation and Memory* (Plenum, New York, 1972).
20. Montarolo, P. G., Castellucci, V. F., Goelet, P., Kandel, E. R. & Schacher, S. *Soc. Neurosci. Abstr.* **11**, 795 (1985).
21. Schacher, S., Montarolo, P. G., Castellucci, V. F., Kandel, E. R. & Goelet, P. *Science* (in the press).
22. Flexner, J. B., Flexner, L. B. & Stellar, E. *Science* **141**, 57-59 (1963).
23. Barondes, S. H. *Short-Term Memory* (eds Deutsch, D. & Deutsch, J. A.) 379-390 (Academic, New York, 1975).
24. Frost, W. N., Castellucci, V. F., Hawkins, R. D. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8266-8269 (1985).
25. Murdoch, G. H., Rosenfeld, M. G. & Evans, R. M. *Science* **218**, 1315-1317 (1982).
26. Waterman, M., Murdoch, G. M., Evans, R. M. & Rosenfeld, M. G. *Science* **229**, 267-269 (1985).
27. de Crombrughe, A., Bushby, B. & Bug, H. *Biological Regulation and Development* Vol. 3B (eds Goldberger, R. F. & Yamamoto, K. R.) 129-167 (Plenum, New York, 1984).
28. Nagamine, Y. & Reich, E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4606-4610 (1985).
29. Yamamoto, K. R. *A. Rev. Genet.* **19**, 209-252 (1985).
30. Ashburner, M. *Heat Shock: From Bacteria to Man* (eds Schlesinger, M. J., Ashburner, M. & Tissieres, A.) 1-9 (Cold Spring Harbor Laboratory, New York, 1982).
31. Murdoch, G. H., Evans, R. M. & Rosenfeld, M. G. in *Biochemical Actions of Hormones* Vol. 12 (ed. Litwack, G.) 38-69 (Academic, 1985).
32. Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1496-1500 (1980).
33. Sabatini, D. D., Kreibich, G., Morimoto, P. & Adesnik, M. *J. Cell Biol.* **92**, 1-22 (1982).
34. Rothman, J. E. *Scient. Am.* **253** (3), 74-89 (1985).
35. Razin, A. & Friedman, J. *Prog. Nucleic Acid Res. molec. Biol.* **25**, 33 (1981).
36. Groudine, M. & Weintraub, H. *Cell* **30**, 131-139 (1982).
37. Cajal, S. R. *Histologie du Système Nerveux de l'Homme et des Vertébrés* Vol. 2 (transl. Azoulay, L.) (Instituto Ramon y Cajal, Madrid, 1955).
38. Kandel, E. R. *Harvey Lect.* **73**, 19-92 (1979).
39. Purves, D. & Lichtmann, J. W. *Principles of Neural Development* (Sinauer, Sunderland, Massachusetts, 1985).
40. Bailey, C. H. & Chen, M. *Science* **220**, 91-93 (1983).
41. Greenough, W. T. in *Neurobiology of Learning and Memory* (eds Lynch, G., McGaugh, J. L. & Weinberger, N. M.) 170-478 (Guilford, London, 1984).
42. Richards, G. & Ashburner, M. *Biological Regulation and Development* Vol. 3B (eds Goldberger, R. F. & Yamamoto, K. R.) 213-253 (Plenum, New York, 1984).
43. Nevins, J. R. *CRC crit. Rev. Biochem.* **12**, 307-322 (1985).
44. Cochran, B. H., Reffel, A. C. & Stiles, C. D. *Cell* **33**, 939-947 (1983).
45. Lau, L. F. & Nathans, D. *EMBO J.* **4**(12), 3145-3151 (1985).
46. Greenberg, M. E., Green, L. A. & Ziff, E. B. *J. biol. Chem.* **260**, 14107-14110 (1985).
47. Krüger, W., Schubert, D. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7330-7334 (1985).
48. Morgan, J. I. & Curran, T. *Nature* (in the press).
49. Weinberg, R. A. *Science* **230**, 770-776 (1985).
50. Lewin, R. *Science* **233**, 159 (1986).
51. Treisman, R. *Cell* **42**, 889-902 (1985).
52. Blanchard, J.-M. *et al. Nature* **317**, 443-445 (1985).

ARTICLES

Petrological and tectonic segmentation of the East Pacific Rise, 5°30'–14°30' N

Charles H. Langmuir*, John F. Bender† & Rodey Batiza‡

* Lamont-Doherty Geological Observatory, Palisades, New York 10964, USA

† University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA

‡ Washington University, St Louis, Missouri 63130, USA

Lavas from the fast-spreading East Pacific Rise are geochemically diverse even within a single tectonically defined spreading cell. Within such spreading cells, small offsets of the rise axis are often boundaries between petrologically distinct magmatic units which must be supplied independently from beneath the ocean crust. Volcanics erupted near the small offsets can have chemical characteristics similar to those previously found near transform faults.

IN April 1985, we carried out an extensive dredging program of the East Pacific Rise (EPR) between 5°30' N and 14°30' N. This region of the EPR is fast-spreading (9–11 cm yr⁻¹ full rate) and contains two major transform faults, the Siqueiros at 8°22' N and the Clipperton at 10°20' N (Fig. 1). There are several smaller offsets identified by Macdonald *et al.*¹, which they named overlapping spreading centres (OSCs). Lonsdale^{2,3} identified similar features between 5°30' N and Siqueiros, which he called non-transform offsets. Our program was designed to examine the effects of transforms and smaller offsets on the chemistry of ocean ridge basalts, and through the chemistry to improve our understanding of magmatic processes along ocean ridges. Another aim of the study was to sample a substantial length of a fast-spreading ridge to test ideas about the relationship between spreading rate and the degree of observed mantle heterogeneity^{4–9}.

Previous work on the chemistry of ocean ridge basalts erupted near transform faults suggested that proximity to transforms leads to several petrological effects, including cooler magmatic temperatures^{10–12}, eruption of a wider range of composition^{12–15}, higher-pressure fractionation^{16,17} and slightly smaller extents of melting^{13,16}, which can lead to eruption of 'enriched' basalts when a veined mantle is present. These petrological effects, which collectively can be called a transform fault effect (TFE), make sense in terms of the cooling of lithosphere and asthenosphere that results from truncation of a spreading ridge by an offset^{16,18}. If these various effects are caused by lithospheric and asthenospheric cooling, the magnitude of the effect should correlate with the temperature (and hence, age) of the truncating lithosphere.

Earlier work also showed that there were abrupt changes in major and trace element chemistry across some transform faults

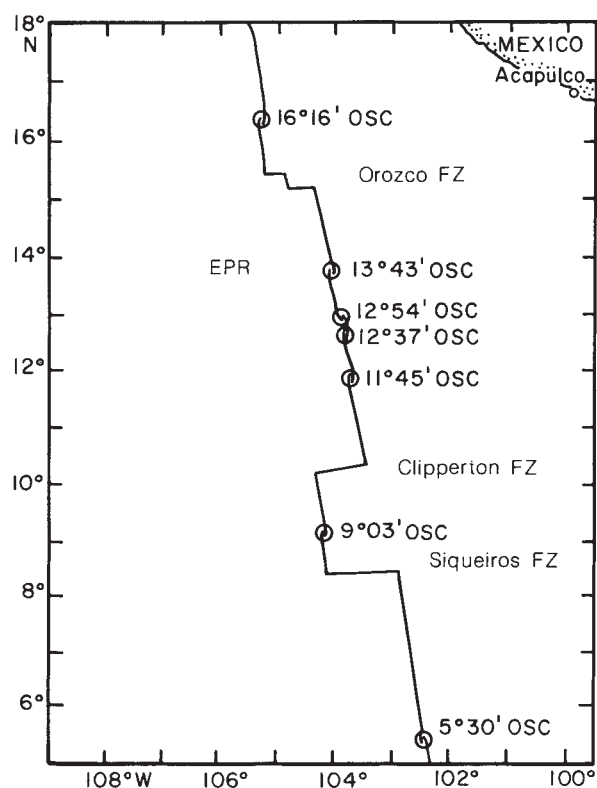


Fig. 1 Map showing general location of CHEPR expedition on the EPR, modified after ref. 1. The major transforms and overlapping spreading centres are indicated.

('transform discontinuities') at slow- and medium-spreading ridges^{16,19,20}. On this basis, it appeared that transforms might separate coherent geochemical units—lengths of ridge crest along which basalts are chemically related by crystal fractionation or different extents of melting¹⁶. This idea, however, was based on sampling immediately adjacent to transforms, and needed to be tested by closely spaced sampling of a series of transform-bounded ridge segments.

The petrological variations associated with spreading centre offsets are just one aspect of a more general problem concerning the magmatic and tectonic segmentation of ocean ridges. Schouten and Klitgord²¹ introduced the useful concept of 'spreading cells' to describe transform-bounded ridge segments, and they showed that, in the Atlantic, spreading cells could be long-lived tectonic units. Francheteau and Ballard²² suggested that large portions of the EPR (~300 km) may be a single cell, although Ballard *et al.*²³ also noted the existence of secondary topographic highs separated by 'relay zones'. It has also been suggested^{1,2} that segmentation of the EPR needed to be considered in the context of both OSCs and transform faults, and thus could be <300 km.

The spreading cell idea has gradually evolved into a general magmatic model for ocean ridges. The centrepiece of this model is that there are magmatic injection centres regularly spaced along the ridge axis which correspond with long-wavelength bathymetric highs^{22,24,26}. Magma moves out from the central injection point along the ridge in a manner analogous to the observed rift eruptions on Iceland²⁵. Such an idea has received support from bathymetric studies²², from the location of hydrothermal fields^{22,26}, from interpretation of multi-channel seismic results²⁷, and from theoretical consideration of potential Rayleigh-Taylor instabilities during melt production beneath ocean ridges^{24,26,28}. In its simplest form, this central injection model predicts that the temperature of eruption of lavas should decrease regularly from a maximum at the bathymetric high at

the centre of a spreading cell to lower temperatures at the boundaries²⁹.

A contrasting, but not mutually exclusive idea, is that magma supply is not confined to the centres of long-wavelength cells, but that there are multiple injection centres distributed along the ridge, with edge effects at offsets which locally lead to cooler temperatures and possibly lower extents of melting (see refs 13–15; ref. 16, Fig. 12). The multiple injection model predicts less systematic long-wavelength variations in temperature and chemistry along axis except in the vicinity of large offsets, where the cold edge effect leads to lower-temperature lavas.

Thompson *et al.*²⁹ recently presented preliminary petrological and photographic results from a short segment of the EPR between 10°20' and 11°45' N. They found at least five different basalt types, and suggested that they were separated into two discrete geographical units with edges at the Clipperton transform and the two OSCs identified by Macdonald *et al.*¹. They also suggested that their data strongly supported the central injection model, and that each bathymetric high between OSCs or transforms was a single injection centre, away from which magmas became progressively more fractionated because of cooling during passage down axis. Inspection of their figures, however, reveals that the data show a narrow range in degree of fractionation ($(\text{Mg}/\text{Mg} + \text{Fe}) = 54 \pm 3(1\sigma)$) except for the dredge locations within 30 km of the Clipperton transform fault. In addition, there are at least two distinct magma types on each bathymetric high, which could be interpreted as evidence against the central injection model. Thus their data do not clearly discriminate between the central injection and multiple injection hypotheses.

Our dredging program was designed to recover samples on the appropriate scale to address the petrological significance of transform faults and smaller ridge offsets, and to discriminate between central injection and multiple injection models of ocean crust formation. The results and discussion which follow are the first-order results based on ~250 shipboard analyses of glass chips, supplemented by ~150 new major and trace element analyses on shore of hand-picked glass powders and 50 analyses of rare-earth elements (REE) by isotope dilution and neutron activation. A fuller evaluation based on data from chemical, mineralogical and isotopic studies will be given elsewhere.

Results

During the cruise, there were 122 successful dredges along 980 km of the EPR. Most of the dredge sites were located using a 12-kHz wide-beam echo sounder, conventional satellite navigation and large-scale versions of Sea Beam maps provided by Jeff Fox and Peter Lonsdale. All dredges were thoroughly described, and representative samples (~8 per dredge) were cut and had glass removed on board. Dredge tracks were usually made along axis and were usually kept to <1 km. One to four glasses per dredge were chemically analysed on board with a direct-current plasma emission spectrometer. The on-board analyses enabled us to determine where the significant changes in chemistry were occurring and to sample those areas more extensively.

Axial depth and linearity. A longitudinal profile of axial depth in our study area is shown in Fig. 2. The minimum depth was picked approximately every 2' of latitude from large-scale versions of the Sea Beam charts^{1,2}, and from our shipboard echo sounder data where there was no Sea Beam data between 7° N and the Siqueiros fracture zone. Macdonald *et al.*¹ pointed out many of the critical features of axial depth north of the Siqueiros transform. In particular, they noted the correspondence of OSCs with local maxima in axial depth. Although most of the depth profile in Fig. 2 is directly derived from the Macdonald *et al.*¹ maps, it differs from their profile in subtle but important ways. In particular, the Fig. 2 profile plots a single minimum depth except where two overlapping limbs of the EPR are completely unambiguous. The petrological data do not support the existence

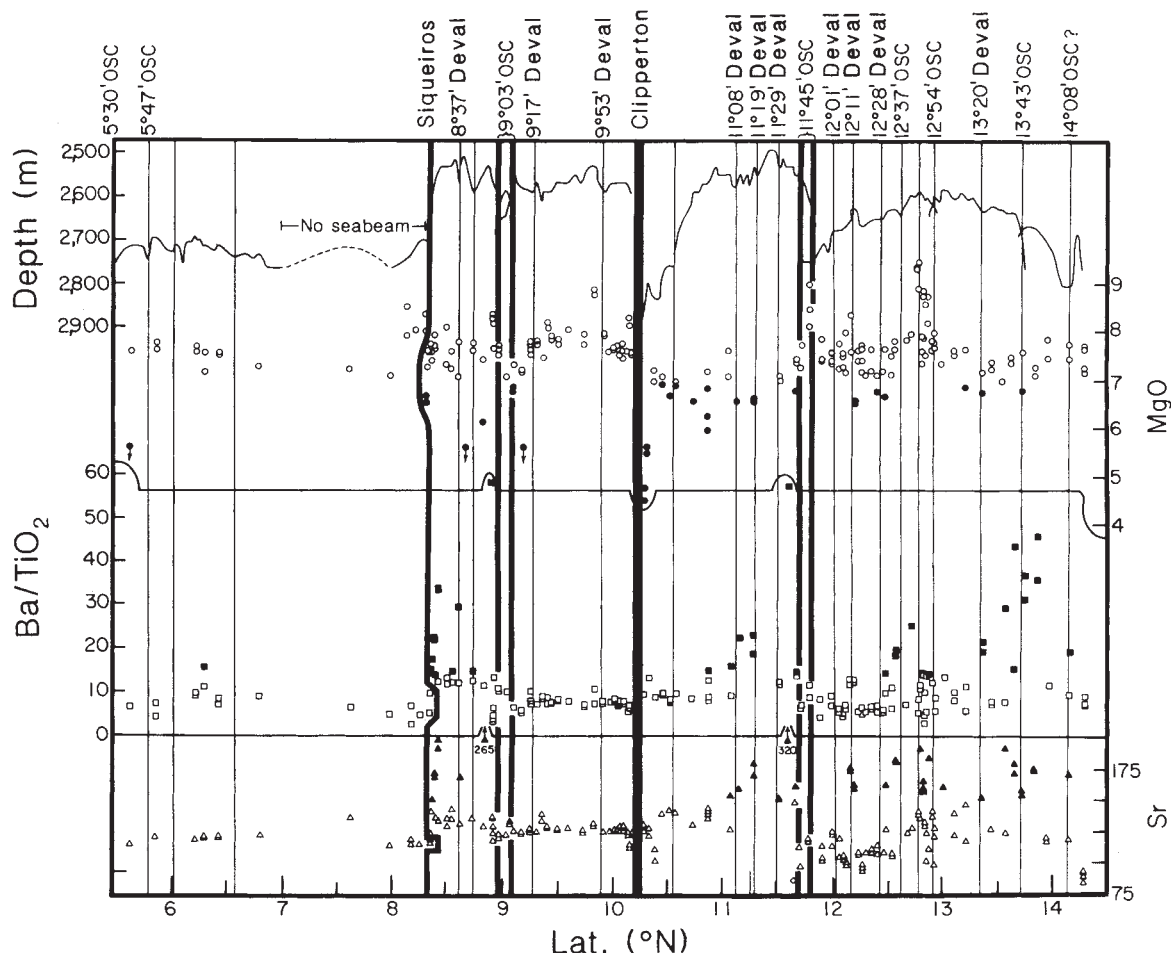


Fig. 2 Plot of axial depth¹⁻³ and various chemical parameters for zero-age basalts versus latitude. Thick vertical lines delineate proposed spreading cell margins at the major OSCs and the two transform faults. Thinner vertical lines occur at small-offset OSCs and 'deviations from axial linearity' (devals) along the EPR. Filled symbols are for samples with MgO <7%, Ba/TiO₂ >14 and Sr >150 p.p.m., respectively. Ba/TiO₂ is a ratio similar to La/Sm. Normal, depleted ocean ridge basalts usually have Ba/TiO₂ ratios of 4-8. Enriched MORB with (La/Sm)_N >1 generally have Ba/TiO₂ >23. Most chemical data were determined at sea by plasma emission spectrometry. Approximately 50 mg of glass chips were run for SiO₂, Al₂O₃, CaO, Na₂O, FeO, MgO, TiO₂, Sr and Ba. Because of approximate weights, sums were normalized to 99.5%. Precision based on 10 replicate analyses of a standard run on board an unknown throughout the cruise was SiO₂, 0.5%; Al₂O₃, 2%; FeO, 1%; MgO, 1.5%; CaO, 2%; Na₂O, 1.5%; TiO₂, 2%; Sr, 2%; Ba, 10%. All data are 1σ in relative per cent (50.0 ± 0.25 for SiO₂). Accuracy, based on samples rerun at Lamont, was worse than precision, and there were slight biases of as much as 3% relative for Na₂O and Sr. Some Ba values at sea were anomalously high, which we attribute to contamination with Mn-oxide rinds. Most high-Ba samples have been rerun at Lamont.

of an OSC at 11°15' N, and that OSC has been omitted.

The four large offsets in the region conveniently divide the EPR into five long-wavelength features, separated by the heavy vertical lines in Fig. 2. Across three of the four large offsets for which depths from both sides are shown there are distinct steps of >100 m in the axial depth. From Clipperton to the north end of the study area the axial depth increases substantially approaching the large offsets, giving a 'domed' shape to the long-wavelength features. South of Clipperton, 'domes' are not obvious, although depths do increase over short distances approaching the 9°03' OSC and the northern side of the Siqueiros transform. Superimposed on all the long-wavelength features are shorter-wavelength (~25 km) undulations. Some of the deep points of these undulations correspond to the small-offset OSCs of Macdonald *et al.*¹, but most do not. For example, just south of Clipperton there are several pronounced local depth anomalies even though there are no small-offset OSCs.

Another way of viewing the EPR is to note the changes in its strike. Lonsdale^{2,3}, for example, noted several kinks in the EPR between 4° and 6°S. Macdonald *et al.*¹ noted a 'regional sinuosity', and Ballard *et al.*²³ noted the existence of short

segments of slightly varying strike separated by what they called 'relay zones'. These changes in strike often occur where there are local depth maxima.

The petrological results presented below suggest that some of the short-wavelength variations in axial strike and depth are magmatically significant. Many of the short-wavelength undulations may be individual small volcanoes. Our initial inclination was to try and preserve a direct connection between the short-wavelength petrologic segmentation of the EPR and the 'spreading cell' concept. However, the short-wavelength structure of the EPR is likely to be temporally variable, and a fundamental feature of the 'spreading cell' idea²¹ is its connection with long-lived tectonic fabric of the ocean floor. Therefore, 'spreading cell' might be an appropriate term for each of the five long-wavelength bathymetric features in this area. Boundaries of spreading cells would be transforms or large-offset OSCs; these boundaries are relatively long-lived and usually have a significant change in axial depth across them. In addition, we suggest that the boundaries of the short-wavelength features within spreading cells be called 'devals' (or 'deviations from axial linearity'). The term 'deval' would include kinks, bends

(which are more gradual changes in strike than kinks) and small ridge offsets with or without overlap. (After the cruise, Batiza and Margolis³⁰ named one type of deval a 'SNOO', or small non-overlapping offset, and proposed a stochastic-spreading genetic model to account for such features.) We introduce the term 'deval' because the boundaries of the short-wavelength features can be petrologically important, yet their morphologic expression is diverse. There is not always an offset or overlap, hence terms which rely on these words are not general enough. Positions of some prominent devals are indicated in Fig. 2. Devals cannot always be identified with certainty on Sea Beam contour maps because of navigational inaccuracies, the nature of the contouring algorithm, and the subjective judgements involved in fitting separate Sea Beam swaths together by hand. Thus we give several examples of the importance of devals but do not present a comprehensive catalogue of their occurrence. **Temperature, axial depth and width.** One aim of the dredging program was to determine if there were systematic changes in the chemistry of basalts with changing depth of the EPR axis. Figure 2 shows the range of MgO contents of 200 glasses from our 100 zero-age dredge hauls. The magnesium content of glasses is directly proportional to liquidus temperature when olivine is close to the liquidus, hence the variations in MgO approximately reflect the relative variations in the quenching temperature of the lavas.

Figure 2 shows that there is no overall correlation between the depth of the rise axis and the MgO content of the erupted lavas. The EPR between Clipperton and the 11°45' OSC is, in general, a shallow segment, and yet it has the lowest mean MgO (that is, lowest temperature of erupted lava), while the adjacent ridge segment between the 9°03' OSC and Clipperton, at approximately the same depth, has the highest mean MgO.

For each spreading cell defined by the major offsets, there is no overall correlation between MgO and distance from the centre of the segment. Some of the highest-MgO lavas occur near the centres of three of the five long-wavelength features, but so do lavas with average MgO contents. High-MgO lavas also occur at some of the margins of these features. Low-MgO, and even high-SiO₂ lavas occur immediately adjacent to the northern side of the Clipperton transform, the 9°03' OSC and the 5°30' OSC, but andesites also occur midway between Siqueiros and the 9°03' OSC. Thus, in some cases, there may be a tendency for the centres of spreading cells to erupt hotter magmas, and the edges to erupt cooler magmas, but the data are noisy.

It is possible that on a shorter wavelength, samples from the centres of some of the secondary highs have higher MgO (for example, 9°50' N, 12°15' N and just south of Siqueiros). In most cases, however, our dredge spacing is not close enough to examine the detailed chemical variations for the secondary highs separated by devals.

In contrast to the complex and irregular relationship between MgO and depth, there may be a correlation between the mean regional MgO content and the mean regional width of the rise axis (Fig. 3). This correlation is not apparent over very short distances and for some unusual areas, such as just south of the 9°03' OSC where the axis is exceptionally broad, but does apply when data are averaged over distances of ~50 km.

Macdonald *et al.*¹ inferred from axial geomorphology that the cross-sectional shape of the rise axis reflected the robustness of the magmatic budget. The width in Fig. 3 is a crude measure of cross-sectional shape, and thus confirms much of Macdonald *et al.*'s¹ discussion. In general, where the EPR is narrow and triangular, such as north of Clipperton, it reflects relative starvation of the magmatic budget. Erupted lavas tend to be cooler and to have low MgO, and show a pronounced edge effect. Where the EPR is broad, such as south of Clipperton, the magma supply is robust, the mean MgO content is higher and there is no perceptible edge effect either in terms of bathymetry or chemistry. The EPR also shows a substantial change in depth approaching Clipperton on the north, but is remarkably constant

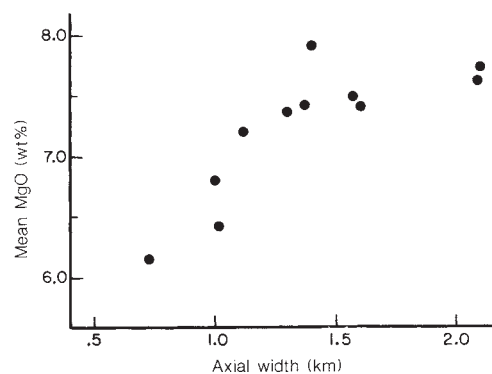


Fig. 3 Width of the EPR axis plotted against mean MgO content of the axis. Width was measured approximately every 2' of latitude as the distance between Sea Beam contours 40 m below the summit, using the maps from refs. 1-3. Standard deviations on the width estimated from this method are large, and the correlation should be regarded as qualitative.

in depth on the south. Thus it may be the slope of the depth along axis rather than the absolute depth that correlates with magmatic temperatures. The bathymetry, shape of the rise axis and chemistry, therefore, do appear to be related, and reflect a balance between magma supply and edge effects. The balance can be highly variable even over short distances, such as across the Clipperton transform, and need not be related to position along a long-wavelength bathymetric feature.

As axial width seems to reflect the magmatic budget, one might expect a correlation between axial width and the presence of hydrothermal activity. It is also intriguing to speculate that the axial shape may reflect the width or depth of axial magma chambers.

Trace element geochemistry. Figure 2 also shows variations in Ba/TiO₂ ratios and Sr content of the basalt glasses. Both these parameters, but especially the Ba/TiO₂ ratio, can be indices of enrichment of the mantle source. Ba/TiO₂ is analogous to and correlates with the better known (La/Sm)_N ratio³¹. Normal ('N-type') MORB (mid-ocean-ridge basalt) glasses from both the Atlantic (~23° N, 54° N) and the Pacific (23° N and 44° N) have Ba/TiO₂ ratios between 2 and 12. Slightly more enriched, or transitional ('T-type') MORB, like those from the FAMOUS area along the Mid-Atlantic Ridge, have Ba/TiO₂ ratios between 25 and 40. As Fig. 2 shows, several discrete lengths of ridge in our study area have Ba/TiO₂ ratios <12. Such N-type MORB occur from the 5°30' OSC to Siqueiros, from the 9°03' OSC north to Clipperton, and for the first 50 km north of Clipperton. From the 11°45' OSC to just south of the 12°37' OSC, 13 out of 15 dredges are N-type. Thus we recovered exclusively normal MORB from about two-thirds of the ridge length surveyed.

Surprisingly, one-third of the length of the EPR in this region contains at least some basalts more enriched than normal MORB. Such basalts occur from Siqueiros to the 9°03' OSC, from a deval at 11°08' to the 11°45' OSC, around the 12°37' OSC, and around the 13°43' OSC. In terms of regional distribution, the portions of the EPR which contain relatively enriched basalts are not systematically distributed with respect to the long-wavelength bathymetric features discussed above. The most enriched basalts in the region occur just south of the large OSCs at 11°45' N and 9°03' N. Some basalts recovered near all of the OSCs in our study region (excluding 11°15') are more enriched than those recovered from the adjoining ridge segments. However, not every sample near OSCs is enriched. Devals also can be the location of relatively enriched basalts. In particular, basalts from the 8°37', 12°11' and 13°20' devals are more enriched than basalts from the axis immediately to the north and south (see Figs 2 and 5).

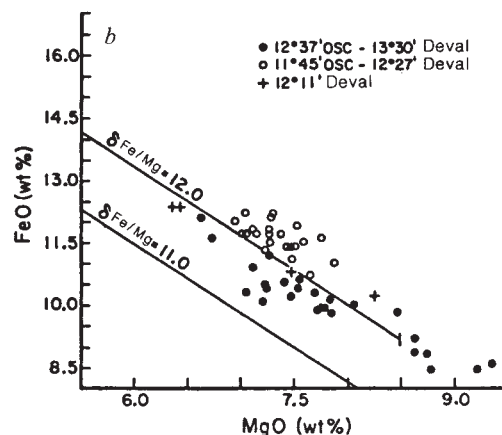
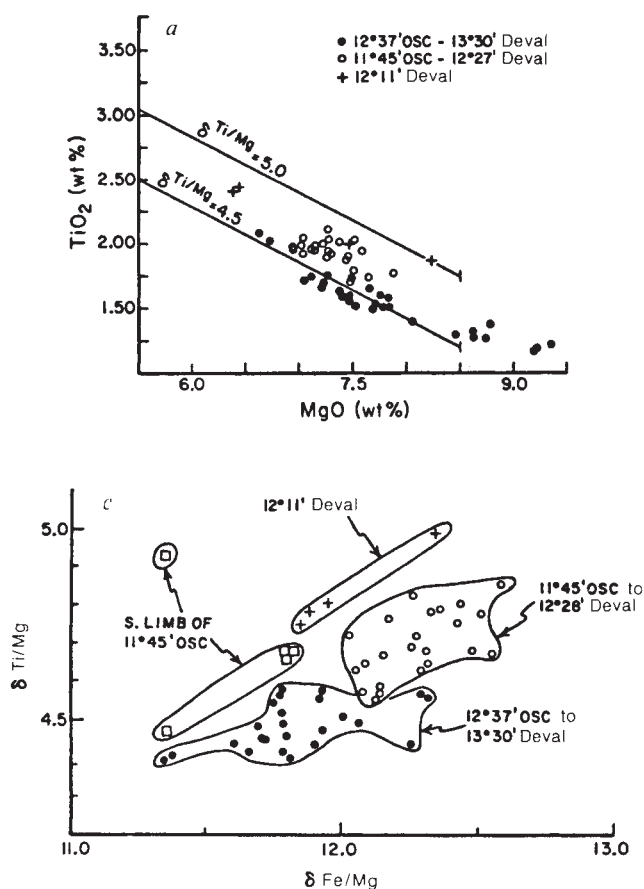


Fig. 4 Demonstration of major element variability along the EPR and the nature of the δ parameters. *a*, *b*, MgO/FeO and MgO/TiO₂ diagrams, all oxides in weight per cent. When the axes are rotated such that x' is parallel to the low-pressure fractionation path, then y' , which is the δ value, reflects differences in the fractionation paths. Lines parallel to the x' -axis at different values of y' ($+\delta$) are shown for illustration. Samples with different δ values cannot be related to one another by low-pressure fractionation. *c*, Plot of δ Fe/Mg versus δ Ti/Mg showing distinct major element characteristics of different ridge segments. Samples from 12° to 12°28' are completely distinct from those recovered between 12°37' to 13°20' N. Samples from near the deval at 12°11' N plot separately, as do samples from south of the 11°45' OSC.

The most notable feature of these data is the large number of occurrences of T-type MORB, and the significant number of dredges which recovered both T-type and N-type MORB. In the North Atlantic, T-type MORB are restricted to regions near hotspots, even where there are large offset transforms^{31,32}. Extensively sampled regions near the Kane fracture zone, for example, have yielded exclusively N-type MORB (ref. 32 and C.H.L. unpublished data). The fact that there are large regions with no enriched MORB, and discrete regions where enriched MORB are found, suggests that the mantle beneath this region of the EPR is not ubiquitously veined with enriched material.

Previously, it has been proposed that fast-spreading ridges are more homogeneous in their chemistry than slow-spreading ridges^{4,5,7}. The data in Fig. 2 demonstrate that basalt chemistry on a substantial portion of the EPR is not homogeneous, but instead is highly variable. Indeed, the mean Ba/TiO₂ and the distribution of enriched material on the EPR between Siqueiros and 14° N is very similar to the ratio that can be inferred from rare-earth data from the South Atlantic, where there are multiple off-axis hotspots^{33,34}. These new data from the EPR are difficult to reconcile with correlations between spreading rate and the extent of geochemical variability which have been previously postulated. The data show that geochemical enrichment is not restricted to off-axis seamounts^{8,35}, but is quite common on axis as well. Thus, models which attempt to explain normal basalts on axis and both normal and enriched basalts off axis^{8,35,36} may need to be re-evaluated. Note, however, that the enriched basalts at zero age are preferentially located near offsets. Thus there may be lower extents of melting of a veined mantle beneath the offsets—suggesting that the offsets are at the margins of mantle melting anomalies. This explanation of the enrichment would be similar to that proposed for seamounts on older crust^{8,35}, but the spatial systematics leading to enriched basalts would occur

parallel rather than perpendicular to the rise axis. The remarkable consequence of this explanation is that even devals may be located at the margins of zones of mantle melting.

Chemical discontinuities across ridge offsets. The Sr and Ba/TiO₂ data in Fig. 2 show several abrupt, as well as some less abrupt, chemical changes across ridge offsets. There are prominent, abrupt chemical changes across the Siqueiros transform and the 11°45' OSC. The 9°03' OSC and 12°37' OSC also bound distinct regions in Ba/TiO₂, but there is chemical 'noise' in the immediate vicinity of the boundary. A deval at 11°08' divides discrete chemical regions as well. (Thompson *et al.*²⁹ suggested the 11°15' OSC was a boundary, but we have two dredges, one at 11°10' and the other at 11°20', which have similar chemistry, suggesting that the actual boundary is south of 11°10'.)

The marked discontinuities in trace elements are also present in major elements—and the major elements reveal many more subtle discontinuities. Major elements are strongly affected by crystal fractionation, and usually suites of samples are required to define a regional trend, which can then be compared to another large region (see ref. 31). An example of such an approach is shown in Fig. 4*a*, *b*, where glasses from north of the 11°45' OSC to south of the 12°27' kink are compared with glasses from north of the 12°37' OSC to the 13°20' N bend. These illustrate a gross chemical distinction between these two regions, but it is useful to consider more detailed variations along axis which are not readily apparent from a presentation such as Fig. 4*a*, *b*. This can be done by rotating the axes so that the new reference frame has its x' -axis approximately parallel to the liquid line of descent after plagioclase has joined olivine on the liquidus, that is, for samples with $\approx 8.5\%$ MgO. The values along the rotated y' -axis are given the δ notation because they reflect differences in the characteristic levels of Fe, Ti or Na of the various liquid lines of descent.

Figure 4c illustrates with major elements the discontinuities across the 11°45' OSC and the 12°37' OSC. The dredges from near the 12°11' deval are distinct from the other data in the 11°45' to 12°37' ridge segment. Noting this distinction while at sea, we decided to halve our dredge spacing for this ridge segment to see if we could resolve smaller-scale segmentation. This segment was ideal for such a test, as there was an extended length of the EPR with no OSCs.

Figure 5 shows the results of this smaller-scale study. On the basis of the $\delta\text{Ti}/\text{Mg}$ parameter, this 50-km ridge segment appears to consist of three distinct chemical populations which are not related to one another by low-pressure fractionation processes. (One of the two sample groups from a single dredge appears to violate the regularity.) A small volcanic edifice with a relief of ~50 m and length of ~10 km (see Fig. 5) appears to consist of material with higher Ba/TiO_2 and $\delta\text{Ti}/\text{Mg}$ than the EPR immediately to the north and south. This small volcano is centred on the 12°11' deval.

This small scale of petrological segmentation is not an isolated phenomenon. Figure 6 shows data for the ridge segment between Clipperton and the 9°17' deval. This ridge segment is more homogenous in general than the segments which contain enriched basalts. There is a deval near 9°53' N. Basalts from south of this feature are distinguishable from basalts north of it both in major and rare-earth elements. In both cases, each group is quite homogeneous, except right at the edges of the segments. Thus, in this region a deval separates two discrete petrological units.

Occurrence of high-silica glasses. Glasses ranging in composition from basaltic andesite with 53% silica to dacite with almost 70% silica were recovered in three on-axis dredges and one off-axis dredge. Of these four dredges, three of them were near the major offsets—just north of the 9°03' OSC, just north of the 5°30' OSC and just north of Clipperton. Previously, high- SiO_2 glasses from the ocean floor have been recovered primarily from behind propagating rifts^{14,37}. Although other models which would lead to high- SiO_2 lavas are feasible, the occurrences on the northern limbs of these large-offset OSCs may imply that these limbs are propagating southwards. (On the basis of bottom photography, Thompson *et al.*²⁹ also suggested that the east limb of the 11°45' OSC may also be propagating southward and Hekinian *et al.*⁴⁰ suggested that the western limb of the 12°54' OSC may be propagating southward.) The fourth occurrence of high- SiO_2 glass is even more intriguing as it occurs 40 km north of the Siqueiros transform fault (Fig. 6d). It is not close to a major offset, but occurs just north of the deval at 8°37' N (Fig. 6d). South of the deval are two dredges with normal basalts with no anomalous enrichment. At the deval there is slightly enriched basalt with higher light-rare-earth concentrations and Ba/TiO_2 ratio. The andesite occurrence is exactly analogous to the other occurrences of high-silica lavas (it is on the northern side of a right lateral offset), except that the offset is a deval rather than an OSC or transform.

Discussion

Recent studies have shown that there are several chemical effects which are often associated with transform faults; the results presented above suggest that chemical variations around some OSCs have many of these same characteristics. OSCs serve as the locus for the eruption of enriched basalts; more primitive, more differentiated and even silicic volcanics occur in their vicinity; and they serve as boundaries between chemically distinct ridge segments.

What is even more remarkable is that many of the chemical effects characteristic of transforms can also be present at very small bathymetric offsets (devals) which have only recently been able to be resolved as offsets with along-axis Sea Beam data^{1,2,23}.

The EPR thus seems to be magmatically segmented into petrological units which can have as boundaries, transforms, OSCs and devals. This segmentation is an order of magnitude

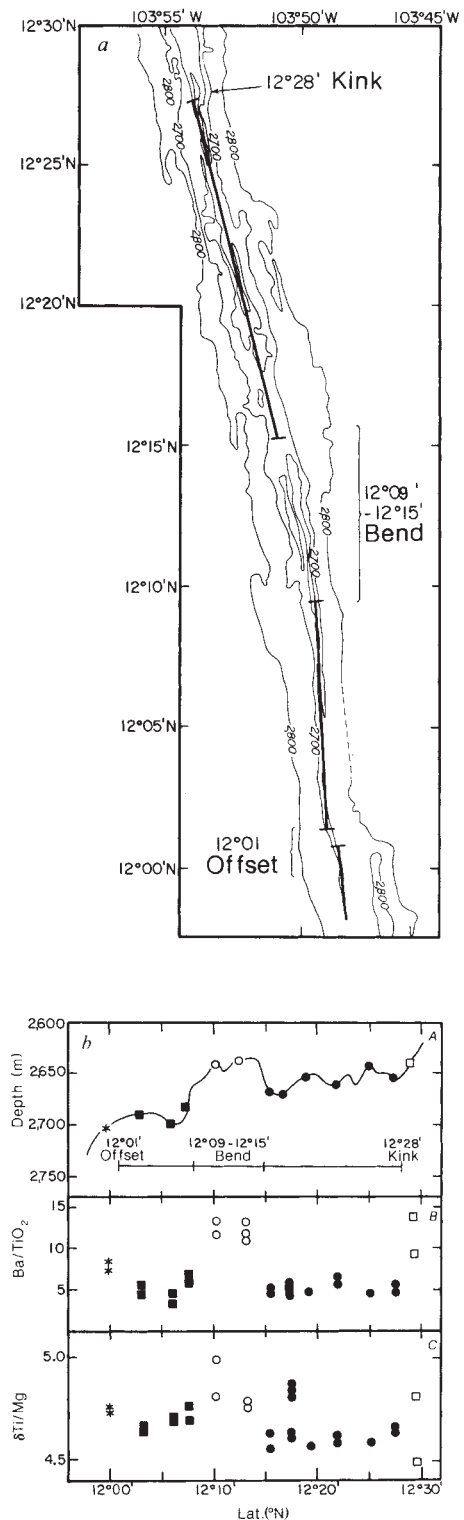


Fig. 5 Documentation of the fine scale of magmatic segmentation which can be present along the EPR. *a*, Sea Beam map¹ of the EPR from 12° to 12°30' N. Locations of deviations from axial linearity (devals) are indicated. *b*, *A*, a depth profile of the axis and the dredge locations. *B*, Ba/TiO_2 ratios for basalt glasses from the dredges. *C*, $\delta\text{Ti}/\text{Mg}$ for the same glasses (see Fig. 4 for explanation of $\delta\text{Ti}/\text{Mg}$). Note that with the exception of one sample group from the dredge at ~12°17' N, the data separate into distinct, regionally coherent groups.

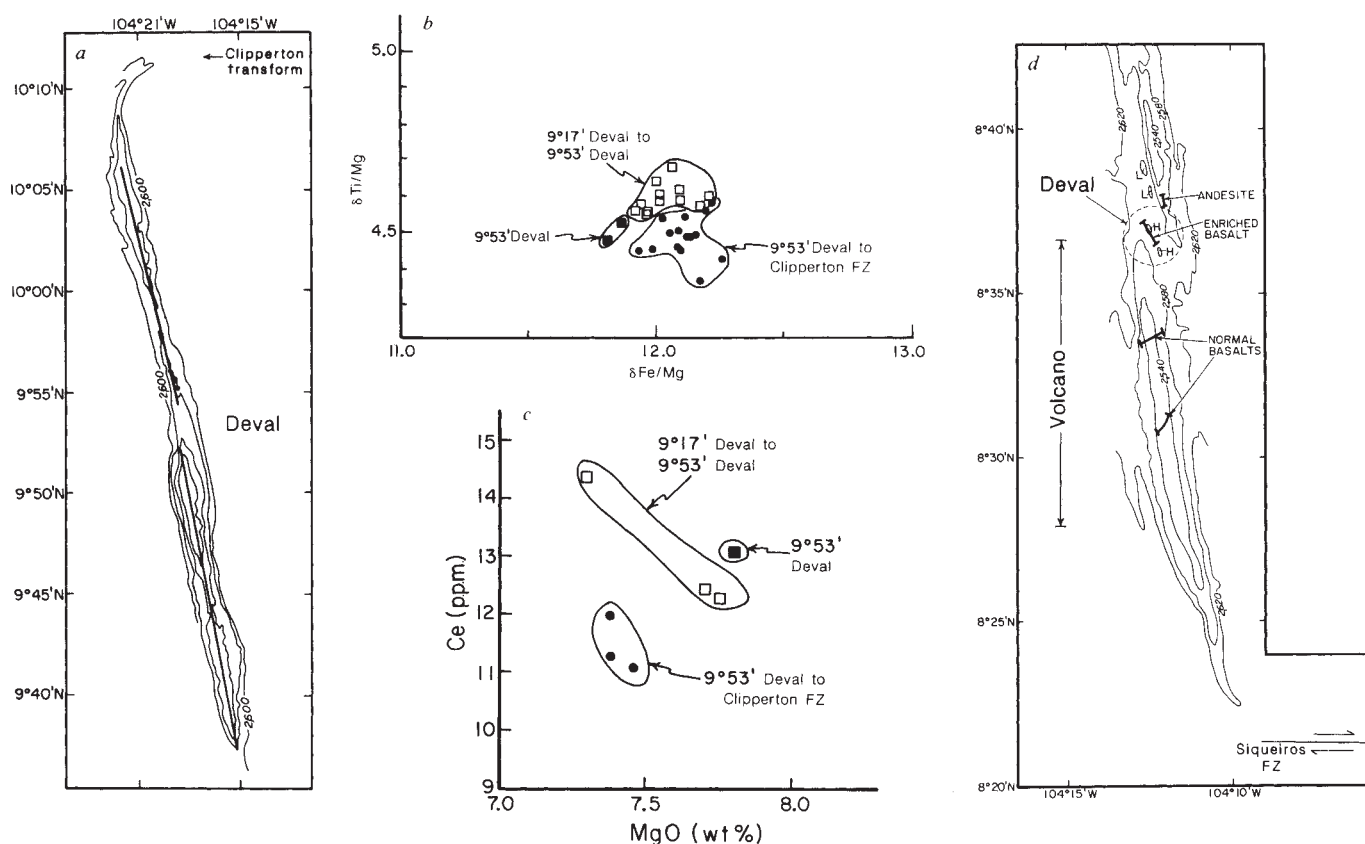


Fig. 6 *a*, Sea Beam map¹ of the EPR just south of the Clipperton transform fault. Note that the EPR seems to consist of a series of sausage-shaped volcanoes. *b*, Plot of $\delta\text{Ti}/\text{Mg}$ versus $\delta\text{Fe}/\text{Mg}$, showing the distinct major element chemistry across the $9^{\circ}53'$ deval. *c*, Ce versus MgO, showing the distinction in rare-earth elements across the $9^{\circ}53'$ Deval. Ce was determined by isotope dilution. *d*, Sausage-shaped volcano north of the Siqueiros transform showing the systematic locations of andesite, normal basalt and enriched basalt.

smaller than long-lived spreading cells. For example, we find that the spreading cell of Francheteau and Ballard²², which includes the whole dome between $11^{\circ}45'$ N and the Orozco fracture zone, is made up of as many as 10 much smaller discrete magmatic segments. The two axial volcanoes of Lonsdale³ between the Gofar and Yaquina transforms may actually be 10 smaller-scale volcanic edifices. Each small segment of rise crest has its own bathymetric expression, and may have its own distinctive chemical signature.

Possible origins of devals. The origin of these features poses an interesting problem which is dependent on the interaction of magma supply with the stress regime at the rise axis, and which must relate in some way to OSCs. There is conceptual continuity across the range from large OSCs with substantial overlap and a deep overlap basin, to small OSCs with almost no decipherable basin, to devals with or without overlaps, and with no basin. This progression naturally suggests a familial relationship among these features and that a general concept may apply to all of them. The petrological data suggest that a starting point for such concepts should be that all the features are boundaries between regions with separate magma supply.

One possibility would be that the size and importance of the various offsets relates to the depth and scale of mantle upwelling. Transforms separate deep upwelling regions, and can remain distinct for up to 10^8 yr (ref. 21). At the other extreme, devals may reflect the last stage of melt segregation, at which small batches of magma ascend from beneath the crust to form axial magma chambers and associated volcanic cones. Individual devals may have a lifetime of $\approx 10^4$ – 10^5 yr. Large-offset OSCs would be intermediate features with intermediate lifetimes.

An important question is whether offsets propagate. Transforms in this region are quite stable and long-lived. The existence

of a volcanic edifice centred on a deval at $12^{\circ}11'$ N leads us to speculate that new magmatic pulses may sometimes nucleate at a deval, and that (at least some) devals may not propagate but may be simply covered up by new volcanic construction. Thus, a deval may mark the edge of a constructional volcanic event, and may provide a nucleation site for the next volcanic pulse.

Implications for OSCs. The discontinuities in chemistry documented above place constraints on models for the origin of OSCs. Both models^{1,2} suggest that OSCs separate regions of independent magma supply, and in this respect our data agree. Lonsdale's model², however, relies on the presence of a large magma chamber beneath OSCs, and in this respect is difficult to reconcile with the petrological data. Macdonald *et al.*¹ suggested magma chambers were continuous beneath small-offset OSCs and that even for large-offset OSCs the magma chamber would become continuous once one OSC linked with another. They suggested that OSCs develop when a magma pulse interacts with this sinuosity of the ridge axis, and that the sinuosity results from previous OSCs. Thus, in their model OSCs are intimately connected with axial magma chambers. Our data suggest that OSCs, particularly those of large offset, may result from being at the margins of deep upwelling zones with fairly long wavelengths—the relationship with magma chambers is not at all clear. Between such large-offset OSCs there may be multiple devals which never were and never will become OSCs. Thus, large-offset OSCs and devals may reflect different stages, scales and depths of magma supply.

Note that there may be an asymmetry to the types of rocks recovered around the large-offset OSCs. All silicic rocks are on northern limbs. The most enriched rocks are on southern limbs, approximately on strike with the northern limbs. To the extent that silicic rocks reflect rift propagation, this suggests that the

northern limbs are propagating southwards. The southern limbs may be stagnating^{2,3}. The stagnating rift may be the locus of seamount formation because it is no longer an active spreading centre and hence eruptions pile up on the crust and form seamounts. Enriched basalt occasionally may occur because the OSCs are at the margins of the melting anomaly in the mantle where the extent of melting is less, and hence small-scale heterogeneities may be sampled. Sinton *et al.*³⁷ have pointed out that absolute plate motions may be important for understanding some aspects of propagating rifts and Schouten *et al.*³⁸ have attempted to relate off-axis seamount trends to absolute plate motions. Intriguingly, if all the northern limbs on this portion of the EPR are propagating southwards, then there may be a relationship with absolute plate motions, which have a northward component in this region.

Axial magma chambers. These petrological data seem to provide tight constraints on the scale of mixing and homogenization of EPR lavas. The small, discrete chemical units along axis, some of which appear to require tapping of heterogeneous mantle, show that the processes of basalt genesis, including magma formation, segregation from the mantle, injection into the crust and eruption onto the ocean floor, do not homogenize EPR lavas. The only way that there could be a continuous axial magma chamber is if it has many separate regions of supply from below the crust and is not well mixed. Basaltic magma chambers in many geochemical models are envisaged as efficient mixing machines which buffer and homogenize magmas³⁹. This geochemical concept of a magma chamber is difficult to reconcile with our data. If, on the other hand, a physical magma chamber does exist, then the magma chamber must be variously fed, incompletely mixed, and hence substantially zoned and laterally heterogeneous in its composition. However, if this were the case, why are most of the petrological data systematically distributed with respect to devals and OSCs? Therefore, while the data do not strictly require the physical absence of a magma chamber, they would be most consistent with magma chambers that are discontinuous, even with single spreading cells.

Conclusions

Systematic dredging of the EPR with an average spacing of 8 km has allowed a definition of the scale of magmatic diversity and segmentation along 700 km of a fast-spreading ridge. The data have definite implications for the two models of ocean crust formation discussed. Central supply of magma to the crust on the scale of spreading cells, even when such cells are defined by OSCs with finite offset, is difficult to reconcile with the data. Instead there seem to be multiple injection zones from the mantle into the crust. This does not mean that there are no petrological regularities with respect to spreading cells. Note that the most pronounced chemical offsets and most pronounced edge effects occur at the spreading cell boundaries, where there are also

discontinuities in axial depth. The long-wavelength anomalies must relate to the temperature structure and mass transfer processes within the underlying mantle. Apparently the mantle is variable on a much smaller scale than the long-wavelength anomalies, and this variability can be preserved through the crust formation process.

The data also have important implications for the effects of transform faults on the petrological evolution of the ocean crust. The south side of the Clipperton transform, where there is a cold edge, apparently has no petrological edge effect, unless it is on the adjoining, older plate. In contrast, there can be pronounced edge effects at devals, where there is no cold edge. Thus, in many cases there are edge effects, but not cold edge effects. This means that the previous emphasis on the cold edge as an explanation of the petrological manifestations associated with transform faults (and possibly propagating rifts) needs to be re-evaluated. Some aspects of the edge effects may be concerned with along-axis systematics of upwelling, melt segregation and distribution in the crust, and magmatic edges may exist even when there are not tectonic cold edges^{13,16} (hence the occurrence of edge effects at devals). The cycle of magma supply to the ocean crust also seems to be able to overwhelm the tectonic cold edge, such as is occurring south of Clipperton. It is not clear from our zero-age data whether the cold edge has a time-averaged effect which only over the short term can be dominated by the magmatic budget.

Finally, the data show that enriched basalts are common at zero age on this portion of the EPR, although they do seem to occur preferentially at offsets. The general concept of fast-spreading ridges being more homogeneous than slow-spreading ridges such as the Mid-Atlantic Ridge is no longer valid.

We thank the other members of the CHEPR team: Emily Klein, Pat Castillo, Jerry Ross, Terri Smith, Dan Miller, Tim O'Hearn, Mike Thatcher, Alex Biesada, Eliesir Duarte and Ron Comer for hard work and dedication. Captain Tom Desjardins and the crew of R/V *New Horizon* were also very helpful. Collaboration with Emily Klein has contributed substantially to this work. Jeff Fox and Peter Lonsdale gave us large-scale versions of Sea Beam maps which were essential for precision dredging. Geoff Thompson, Bill Bryan and Jim Natland provided preliminary data from cruises to the EPR. Conversations with or reviews by Bill Bryan, Dave Christie, Bob Detrick, Dan Fornari, Jeff Fox, Fred Frey, Dick Hey, Ellen Kappel, Jill Karsten, Kim Kastens, Peter Lonsdale, Ken Macdonald, John Mutter, Jim Natland, Mike Perfit, Paul Robinson, Jean-Guy Schilling, Hans Schouten, John Sinton, Geoff Thompson, and Dave Walker were very useful. This work was supported by NSF grant OCE84-11448 to C.H.L. at Lamont-Doherty Geological Observatory and OCE84-09660 to J.F.B. at the University of North Carolina at Charlotte. The Seed Money program at Lamont-Doherty provided part of the funds necessary to carry out the shipboard analyses.

Received 15 October 1985; accepted 3 June 1986.

- Macdonald, K., Sempere, J.-C. & Fox, P. J. *J. geophys. Res.* **89**, 6049-6069 (1984).
- Lonsdale, P. J. *J. geophys. Res.* **88**, 9393-9406 (1983).
- Lonsdale, P. J. *Bull. geol. Soc. Am.* **96**, 313-327 (1985).
- Allègre, C. J., Hamelin, B. & Dupré, B. *Earth planet. Sci. Lett.* **71**, 71-84 (1984).
- Cohen, R. & O'Nions, R. K. *J. Petrol.* **23**, 299-324 (1982).
- Hamelin, B., Dupré, B. & Allègre, C. J. *Earth planet. Sci. Lett.* **67**, 340-350 (1984).
- Batiza, R. *Nature* **309**, 440-441 (1984).
- Zindler, A., Staudigel, H. & Batiza, R. *Earth planet. Sci. Lett.* **70**, 175-195 (1984).
- Macdougall, J. D. & Lugmair, G. *Nature* **313**, 209-211 (1985).
- Hekinian, R. & Thompson, G. *Contr. Miner. Petrol.* **57**, 145-162 (1976).
- Schilling, J. G. & Sigurdsson, H. *Nature* **282**, 370-375 (1979).
- Christie, D. & Sinton, J. *Earth planet. Sci. Lett.* **56**, 321-335 (1981).
- Bender, J., Langmuir, C. & Hanson, G. J. *Petrol.* **25**, 213-254 (1984).
- Perfit, M. & Fornari, J. J. *J. geophys. Res.* **88**, 10551-10572 (1983).
- Natland, J. *Init. Rep. DSDP* **54**, 833-850 (1980).
- Langmuir, C. & Bender, J. *Earth planet. Sci. Lett.* **69**, 107-127 (1984).
- Sinton, J. & Christie, D. *Eos* **65**, 1138 (1985).
- Fox, P. J. & Gallo, D. *Tectonophysics* **104**, 205-247 (1984).
- Schilling, J.-G., Kingsley, R. & Devine, J. J. *J. geophys. Res.* **87**, 5593-5610 (1982).
- Machado, N., Ludden, J. & Brooks, C. *Nature* **295**, 286-288 (1982).
- Schouten, H. & Klitgord, K. *Earth planet. Sci. Lett.* **59**, 255-266 (1982).

- Francheteau, J. & Ballard, R. *Earth planet. Sci. Lett.* **64**, 93-116 (1983).
- Ballard, R., Hekinian, R. & Francheteau, J. *Earth planet. Sci. Lett.* **69**, 176-186 (1984).
- Whitehead, J., Dick, H. J. B. & Schouten, H. *Nature* **312**, 146-148 (1984).
- Sigurdsson, H. & Sparks, S. R. J. *Nature* **274**, 126-130 (1978).
- Crane, K. *Earth planet. Sci. Lett.* **72**, 405-414 (1985).
- Mutter, J. *et al. Geology* **13**, 629-632 (1985).
- Schouten, H., Klitgord, K. & Whitehead, J. *Nature* **317**, 225-229 (1986).
- Thompson, G., Bryan, B., Ballard, R., Hamuro, K. & Melson, W. *Nature* **318**, 429-433 (1985).
- Batiza, R. & Margolis, S. H. *Nature* **320**, 439-441 (1986).
- Schilling, J.-G. *et al. Am. J. Sci.* **283**, 510-586 (1983).
- Bryan, W., Thompson, G. & Ludden, J. J. *J. geophys. Res.* **86**, 11815-11836 (1981).
- Schilling, J. G., Thompson, G., Kingsley, R. & Humphries, S. *Nature* **313**, 187-191 (1985).
- Humphris, S., Thompson, G., Schilling, J.-G. & Kingsley, R. *Geochim. cosmochim. Acta* **49**, 1445-1464 (1985).
- Batiza, R. & Vanko, D. *J. geophys. Res.* **89**, 11235-11260 (1984).
- Sleep, N. J. *J. geophys. Res.* **89**, 10029-10041 (1984).
- Sinton, J., Wilson, D., Christie, D., Hey, R. & Delaney, J. *Earth planet. Sci. Lett.* **62**, 193-207 (1983).
- Schouten, N., Whitehead, J. & Dick, H. *Nature* (in the press).
- O'Hara, M. J. *Nature* **266**, 503-507 (1977).
- Hekinian, R., Augarde, J. M., Francheteau, J., Gente, P., Ryan, W. & Kappel, E. *Mar. geophys. Res.* **7**, 359-377 (1985).

Global temperature variations between 1861 and 1984

P. D. Jones, T. M. L. Wigley & P. B. Wright*

Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Recent homogenized near-surface temperature data over the land and oceans of both hemispheres during the past 130 years are combined to produce the first comprehensive estimates of global mean temperature. The results show little trend in the nineteenth century, marked warming to 1940, relatively steady conditions to the mid-1970s and a subsequent rapid warming. The warmest 3 years have all occurred in the 1980s.

GLOBAL mean surface air temperature is the most commonly used measure of the state of the climate system. When general issues of climatic change are addressed, global mean temperature change is often used as a yardstick; the age of the dinosaurs was warmer than today, the ice ages were colder, and so on. Paradoxically, in the present era of instrumental meteorology, with data coverage far better than at any earlier time, our knowledge of global mean temperature changes is still uncertain. Variations in global mean air temperature are of considerable importance, as they are a measure of the sensitivity of the climate system to external forcing factors such as changes in carbon dioxide concentration, solar output and the frequency of explosive volcanic eruptions. Quantifying the response of the climate to external forcing changes is a major goal of climatology and a prerequisite for predicting future climatic change. As a step towards this goal, we present here the first global synthesis of near-surface temperature measurements over the land and oceans.

Most earlier estimates of global and hemispheric mean temperature (see refs 1, 2) were based solely on data from land-based meteorological stations. Since >70% of the globe is ocean, one might suspect the global representativeness of such estimates, although on long timescales ($\geq$ decades) the thermal coupling between land and ocean should ensure that the land data largely mirror changes occurring over the oceans¹. Recently, data from ships at sea collected for routine weather forecasting purposes, have been compiled by groups in the United Kingdom^{3,4} and the United States^{5,6}, and these data give us the potential to calculate improved estimates of global mean temperature. Apart from our own work⁷, the only previous attempt to analyse both land and marine data is that of Paltridge and Woodruff^{8,9}. These authors, however, failed to account for inhomogeneities in the marine data, which are substantial (see below and also refs 4 and 10). The quality and coverage of the land data they used was also less than adequate, but this is understandable because they were primarily interested in sea-surface temperature variations.

The land data we use are those from refs 11, 12. These have been carefully examined to detect and correct for non-climatic errors that may result from station shifts or instrument changes, changes in the methods used for calculating means, urban warming, and so on. Although problems still exist^{13,14}, the quality of these data is much better than that of material used in earlier studies. Area averages based on these data show medium to long timescale trends ($\geq$ 10 yr) whose spatial consistency provides a strong pointer to the data's overall reliability^{1,11,12}. The marine data we employ are those in the COADS (Comprehensive Ocean Atmosphere Data Set) compilation⁶ which extends to 1979, and data from the Climate Analysis Center, NOAA,

for 1980–84. We use both sea surface temperatures (SST) and marine air temperatures (MAT).

Marine data problems

Both SST and MAT data contain 'inhomogeneities', variations resulting from non-climatic factors^{4,10,15}. For example, early SSTs were measured using water collected in uninsulated, canvas buckets, while more recent data come either from insulated bucket or cooling water intake measurements, with the latter considered to be 0.3–0.7 °C warmer than uninsulated bucket measurements¹⁰. For marine air temperatures, changes in the size and speed of ships, especially those increases associated with the sail to steam transition, are both thought to have influenced data homogeneity. In addition, many early air temperature observations were not taken in screened locations. Because of these non-climatic factors, both SST and MAT data must be corrected (or 'homogenized') to remove their effects.

Folland *et al.*^{4,16} and Folland and Kates¹⁷, using the UK Meteorological Office (UKMO) data bank³, attempted to overcome these problems by identifying specific sources of error, attempting to quantify these and using this information to make corrections to the raw gridded data. Such corrections have inherent uncertainties because of difficulties in their *a priori* quantification and a lack of knowledge of how most measurements were taken. Information on whether bucket or intake measurements were made has, in most cases, apparently been lost or never recorded. It has also been shown¹⁸ that supposedly homogeneous (that is bucket-only or intake-only) SST data series appear to have non-climatic changes that are similar to those found in mixed data series, suggesting that all historical data sets contain a mix of measurement types. Since 1945, however, it is generally assumed that available SST data contain a reasonably consistent mix of intake and bucket measurements¹⁸.

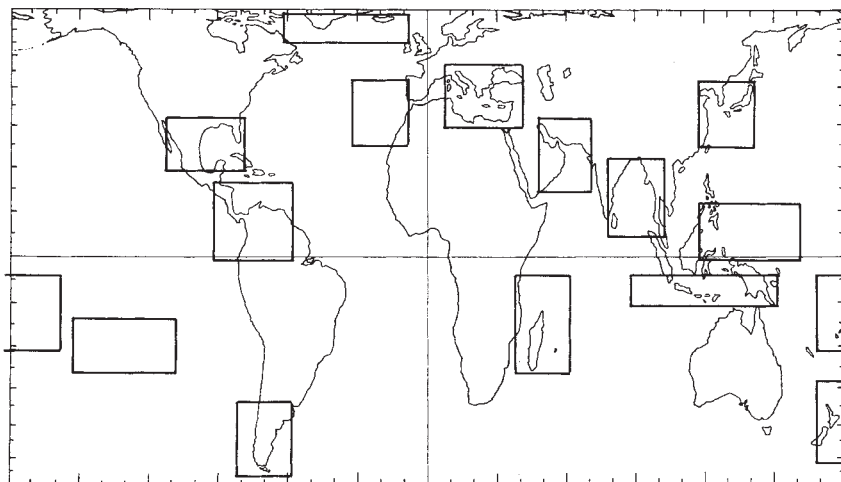
The Folland *et al.*⁴ corrected MAT and SST series have been compared with averages of land-based data by Jones *et al.*^{11,12}. Agreement is reasonable since the start of the twentieth century, although MAT values for the years 1942–45 appear to be too warm in both hemispheres. Before 1900, the marine and land series diverge markedly, with both marine series being about 0.3 °C warmer than the land data.

Correcting the COADS data

The COADS compilation contains some 63.25 million non-duplicated SST observations, of which 0.96 million have been 'trimmed' to remove extreme outliers⁵. While these are more data than in the UKMO SST set (which has about 46 million non-duplicated observations⁴), the effective area and density of coverage is very similar in both data sets. However, unlike the UKMO data set used by Folland *et al.*⁴, none of the data in COADS have been corrected for non-climatic effects. Our first task, therefore, was to homogenize the COADS data. We did

* Present address: Max-Planck-Institut für Meteorologie, Bundesstrasse 55, 2000 Hamburg 13, FRG.

Fig. 1 Map showing the 15 regions where marine air temperatures and land-based temperatures were compared (Peters equal-area rectangular projection).



this by comparing marine and land data in areas where the two abut or overlap (coastal areas and around ocean islands).

The trimmed COADS data include monthly means and medians on a $2^\circ \times 2^\circ$ grid, together with the number of observations in a month and the mean observation date. We compressed the data onto a more manageable grid ($5^\circ \times 5^\circ$ for MAT, $4^\circ \times 10^\circ$ for SST) after first eliminating values where the number and distribution of observations was likely to have produced unrepresentative monthly means, and expressed the values as anomalies from a 1950–79 reference period. As a test of data quality at this stage we calculated hemispheric mean values by appropriately weighting the gridded MAT and SST data (NH, Northern Hemisphere; SH, Southern Hemisphere). Year-to-year variations for these uncorrected data were found to be in excellent agreement with the UKMO corrected data (NHSST, $r = 0.86$; NHMAT, $r = 0.87$; SHSST, $r = 0.88$; SHMAT, $r = 0.75$ over 1856–1979: correlation coefficients calculated using residuals from a 10-yr gaussian filter), but, as expected, the long-term (≥ 10 yr) fluctuations showed marked differences. Similar high frequency correlations between SST and MAT for the uncorrected COADS data (NH, $r = 0.91$; SH, $r = 0.89$) were higher than in the corrected UKMO data (NH, $r = 0.81$; SH, $r = 0.80$).

Because of the high SST–MAT correlation (see also ref. 19), SST data can be corrected by comparison with MAT data, once the latter have been corrected. For the MAT data, any attempt to assess, *a priori*, the magnitudes of errors arising from instrumental changes, changes in observation methods, and the effects of changes in ships' thermal inertia, speed and size (the latter determines the height at which observations were taken), must be fraught with uncertainty. Data reliability and long-term homogeneity can be far more convincingly demonstrated for the gridded land data than for the marine data because land station data homogeneities can be more easily identified, explained and corrected^{11,12}. We therefore use these data directly to correct the marine data. Fifteen regions (see Fig. 1) were chosen in which land and marine data are in close proximity. Area averages of annual mean MAT and land air temperature were calculated for each region using the uncorrected COADS data and the homogenized land data produced by Jones *et al.*^{11,12}. No attempt was made to consider night-time observations only, as used by Folland *et al.*⁴. In addition to the 15 pairs of area averages, annual mean coastal land time series were produced for both hemispheres and compared with the uncorrected hemispheric-mean MAT series.

The 17 land minus MAT time series were then examined for systematic differences between the land and marine data. For the period 1861–1979 (both marine and Southern Hemisphere land data are unrepresentative before 1861 because of poor data

coverage), five distinct periods could be discerned in all 15 regional land minus MAT time series and in the two hemispheric land minus MAT time series. The latter are shown in Fig. 2. The three main periods are: the period up to the 1880s when the MAT data appear to be too warm by 0.4–0.5 °C; the period from the 1900s to 1941 when the MAT data are too cold by 0.1–0.2 °C; and 1946–79 when there is no obvious bias. There is a strong upward trend in the land-minus-MAT difference between the mid 1880s and the late 1900s, and the war years, 1942–45, are marked by anomalously warm MAT values. The consistency between the hemispheres is clear from Fig. 2, and the land minus MAT data for the individual smaller regions, although showing greater inter-annual variability, all show the same features.

The nineteenth century land minus MAT data also show differences between the values before and after about 1873 (see Fig. 2). By examining land, MAT and SST data it can be shown that this difference is also likely to reflect a non-climatic

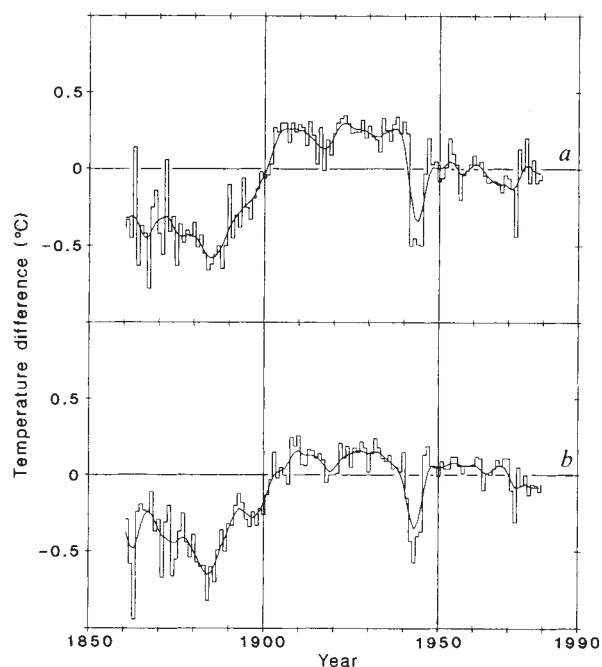


Fig. 2 Temperature differences: coastal land values minus uncorrected COADS marine air temperature values for the Northern (a) and Southern (b) Hemispheres. Smooth curves show 10-yr gaussian filtered values, padded at each end as described in ref. 11.

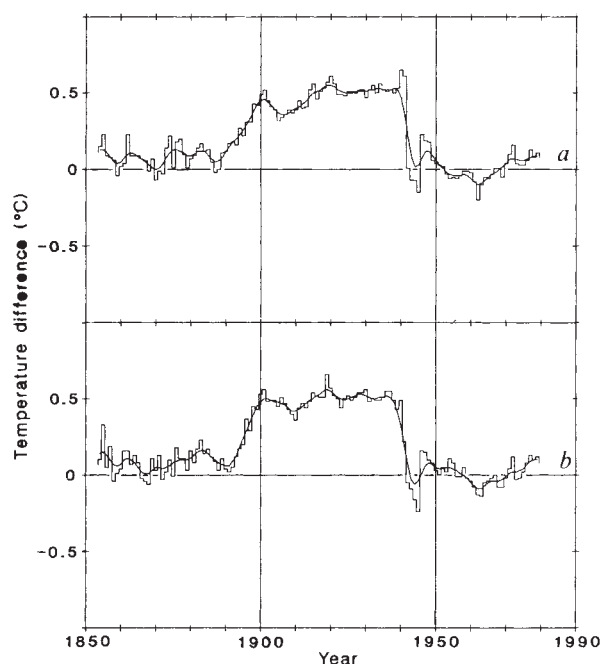


Fig. 3 Temperature differences: corrected marine air temperatures minus uncorrected sea surface temperatures for the Northern (a) and Southern (b) Hemispheres. Smooth curves show 10-yr gaussian filtered values.

inhomogeneity in either the MAT data or the land data, probably the former.

The means and standard deviations of the land minus MAT values are shown in Table 1. The consistency of these values strongly suggests that these land/MAT discrepancies are not climatic in origin. They may, therefore, be used to estimate annual correction factors for the MAT data in order to make these data compatible with the existing homogenized land data. Except for the 1942–45 period, when war conditions apparently prompted observers to measure temperature in unconventional locations⁴, the specific reasons for these non-climatic MAT fluctuations are not known. Although their reality cannot be questioned, there is clearly some uncertainty in the magnitude of the implied corrections.

The correction values we have used (added to the raw MAT data) are (°C): 1861–73, –0.40; 1874–89, –0.48; 1903–41, 0.17; 1942–45, –0.54; 1946–79, 0.0; with linear interpolation between

Table 1 Comparison between coastal land and MAT data

		1861–73	1874–89	1903–41	1942–45	1946–79
NH	$\bar{X}$	–0.35	–0.50	0.23	–0.49	–0.02
	s	0.26	0.11	0.09	0.02	0.12
SH	$\bar{X}$	–0.36	–0.53	0.10	–0.44	0.03
	s	0.23	0.14	0.09	0.09	0.10
NH (9 region average)	$\bar{X}$	–0.36	–0.42	0.17	–0.54	–0.03
	$s(\bar{X})$	0.40	0.21	0.10	0.10	0.05
SH (6 region average)	$\bar{X}$	–0.61	–0.52	0.17	–0.44	0.05
	$s(\bar{X})$	0.57	0.36	0.22	0.15	0.08
Correction		–0.40	–0.48	0.17	–0.54	0.00

$\bar{X}$ = mean land minus MAT value; s = corresponding standard deviation defined by $s^2 = (Y - 1)^{-1} \sum (X_j - \bar{X})^2$ where X_j is the value in year j , and Y is the number of years; $s(\bar{X})$ = standard deviation of the means defined by $(s(\bar{X}))^2 = (n - 1)^{-1} \sum (\bar{X}_i - \bar{\bar{X}})^2$ where n is the number of regions (6 or 9), $\bar{X}_i$ is the mean for region i and $\bar{\bar{X}}$ is the average value of $\bar{X}_i$. The last line shows the inferred correction which was added to the uncorrected annual MAT data.

1889 and 1903. Slightly different corrections were judged necessary for Southern Hemisphere data between 1941 and 1945: 1941, –0.14; 1942–45, –0.44. Most of the transition dates for these correction factors, which are based on a number of considerations, could be altered slightly with no appreciable effect on the resulting corrected MAT values. Although the 0.08 °C difference in the MAT corrections before and after 1873 may be inappropriate if it arises from a land data inhomogeneity, we judge this to be unlikely. It has the effect of slightly reducing the magnitude of the long-term MAT warming between the period before 1873 and today. The corrections generally reflect the mean land minus MAT values shown in Table 1, but the precise values used and the transition dates also take MAT–SST comparisons into account. Our corrections differ markedly from those applied by Folland *et al.*⁴ to their night-time MAT data. This is a clear indication of incompatibilities between the corrected UKMO MAT data and the homogenized land data (see also refs 11 and 12).

Having corrected the MAT data, we can now estimate the SST corrections required to ensure overall compatibility between the land, MAT and SST data by comparing the corrected MAT and raw SST values. Table 2 and Fig. 3 show the hemispheric mean differences between the corrected MAT data and the raw SST data. As with the MAT analysis, three distinct periods can

Table 2 Comparison between corrected MAT data and uncorrected SST data

		1861–89	1903–41	1942–45	1946–79
NH	$\bar{X}$	0.08	0.49	–0.07	0.02
	s	0.08	0.08	0.07	0.09
SH	$\bar{X}$	0.07	0.50	–0.14	0.02
	s	0.04	0.05	0.08	0.08
Correction		0.08	0.49	–0.10	0.00

$\bar{X}$ = mean MAT minus SST value; s = corresponding standard deviation. The correction is the number added to the uncorrected annual SST data.

be discerned: pre-1890 when the SST data are slightly but consistently cooler than the MAT data; 1903–41 when SSTs are markedly cooler than MATs; and post-1945 when there is no consistent difference. Rather complex transitions exist between these three phases. The MAT–SST difference curves are essentially the same in both hemispheres. This is a strong indication that the differences reflect non-climatic effects, and it provides a valuable consistency check on the MAT corrections.

The implied SST corrections, are (°C): 1861–89, 0.08; 1903–41, 0.49; 1942–45, –0.10; 1946–79, 0.0; with linear interpolation between 1889 and 1903. For 1941 we applied a slightly different correction in the Southern Hemisphere, 0.19 °C. As for MAT, these corrections also differ somewhat from those used by Folland *et al.*⁴. In their analysis, SST values were adjusted to ensure compatibility with corrected MAT values, just as we have done. However, since their corrected MAT values must differ noticeably from those produced here, differences in the SST corrections will, in part, reflect these MAT differences.

In our analysis, the difference between the twentieth century SST correction factor before 1941 and after 1946 is 0.49 °C. This difference is in the range (0.3–0.7 °C) generally accepted for the difference between uninsulated bucket and intake SST measurements^{18,20,21}. The precise reasons for the differences that we obtain between the nineteenth century and early twentieth century MAT and SST corrections are uncertain. For MAT, the change is likely to be related to the transition from sail to steam. Between 1880 and 1910, the percentage of steamship tonnage as a fraction of total shipping tonnage rose from ~25 to 75% (ref. 22). Noticeable increases in ship speed occurred over the period 1880–1900, and in ship size over the period 1890–1910

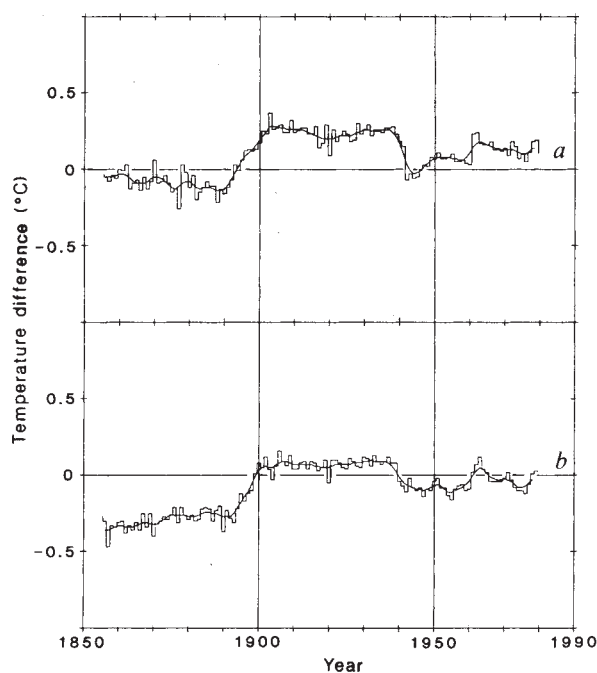


Fig. 4 Differences between the hemispheric-mean sea surface temperature values produced in the present work and those of Folland *et al.*⁴; Northern Hemisphere (a), Southern Hemisphere (b). Smooth curves show 10-yr gaussian filtered values. The implied warmth of the Folland *et al.* SH data relative to the NH (by $\sim 0.2^\circ\text{C}$), is due to their use of 1951–60 as a reference period. Conditions during this decade differed noticeably from the mean conditions during the reference period, 1950–79, used here (see Fig. 5).

(ref. 22). These dates should be compared with the duration of the rising trends in land minus uncorrected-MAT data in both hemispheres shown in Fig. 2. Changes in MAT may be related to exposure changes attendant on the above, and to other changes in instrument exposure procedure which occurred over the same period. For SST, the main reasons for the change may be the standardization of the measuring technique and the introduction of more reliable instruments²³. It is also possible that, in the mid to late nineteenth century, many bucket temperatures were not taken in the shade²⁴. In addition, some of the earlier measurements may have been made with wooden rather than canvas buckets. The latter, being uninsulated and subject to evaporative cooling, produce lower temperature readings.

The overall differences between the hemispheric mean SST values produced here and those of Folland *et al.*⁴ are shown in Fig. 4. The results for an MAT comparison are similar. The discrepancies are large and comparable in magnitude to either set of corrections. The reasons for these differences stem mainly from the different correction factors applied to what are essentially similar raw data. Because there are several sources of data inhomogeneity, we have not attempted to correct for these individually. The result should be more complete than Folland *et al.*⁴ who attempted to make specific corrections for identified sources of inhomogeneity based on physical arguments. Our corrections synthesize the effects of several different factors. However, while they ensure compatibility between the marine and land data, the fact that the reasons for these corrections are uncertain must point towards some remaining uncertainty in our corrected marine data, especially in the nineteenth century.

Global mean temperatures

It is a relatively simple matter to produce estimates of annual global mean surface air temperature using the available (correc-

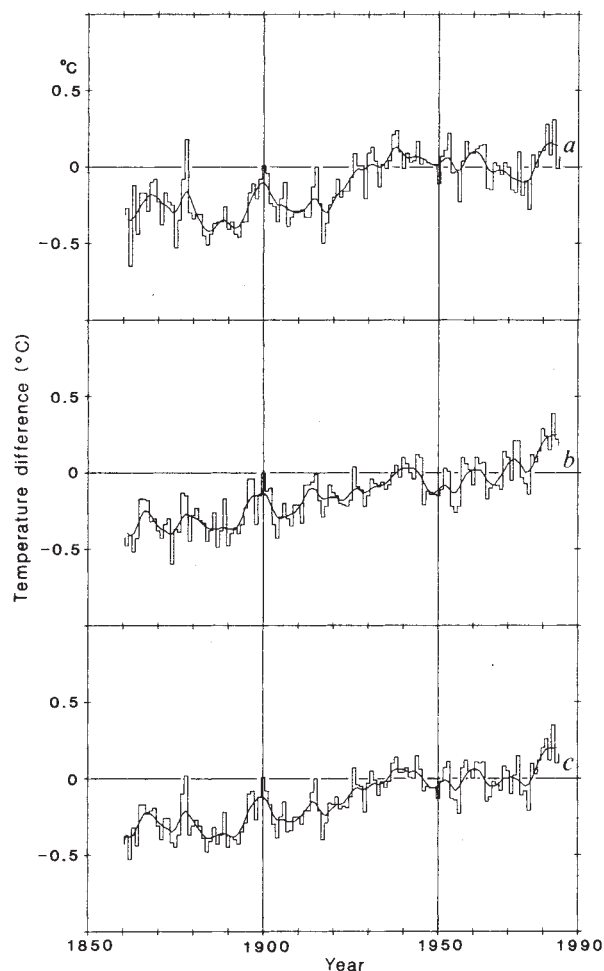


Fig. 5 Global (c) and hemispheric (Northern, a; Southern, b) annual mean temperature variations since 1861, based on sea-surface temperature data to represent the marine domain and using weights corresponding approximately to the maximum coverage for the four domains (method two in the text). Smooth curves show 10-yr gaussian filtered values. 1980–84 values are based on SST data obtained from the Climate Analysis Center, U.S. National Oceanic and Atmospheric Administration (see ref. 29 for information about this data source). These data were adjusted to be compatible with the values in earlier years by comparing values in both hemispheres over the overlap period, 1970–79. The CAC data correlate highly with the COADS data ($r=0.984$ for the Northern Hemisphere mean and $r=0.991$ for the Southern Hemisphere mean).

ted) marine data and the most recent compilations of land data^{11,12}. There are three different ways in which global or hemispheric (land plus marine) averages can be calculated. The first method is to average only those grid point values (with appropriate cosine weighting) for which data exist. This is the way hemispheric means have been produced for the land data^{11,12}. The second and third methods assume that each of the four independent time series (NH and SH land and NH and SH marine, either SST or MAT) are, at all times, representative either of their maximum coverage or of the total areas of the four domains. The results obtained differ but little, and the use of either SST or MAT to represent the marine domains produces only minor differences. We therefore show only results using the second method based on SST data, obtained using

$$T_{\text{global}} = 0.25\text{NH land} + 0.25\text{NH SST} + 0.2\text{SH land} + 0.3\text{SH SST}$$

(Fig. 5) where, after 1957, SH land includes Antarctic data from

Raper *et al.*²⁵, updated. The insensitivity to the precise method of weighting arises because all time series are quite strongly correlated.

The reliability of the time series given in Fig. 5 as true hemispheric and global averages can be questioned because the spatial coverage, even at best, is less than 75% and because the coverage changes with time. Coverage is always much better in the Northern Hemisphere. Coverage before 1900 is generally less than one third of the globe, down to <20% in the 1860s. The question of representativeness of the land data has been considered in detail in refs 1, 11 and 12. Although marine coverage before 1900 is sparse, the spatial correlation length over the oceans is large and limited coverage should still give results representative of a much larger area. Nevertheless, there are large parts of the Southern Hemisphere that nearly always lack data, especially the southern oceans south of 45°S and the whole of the southeastern Pacific (except near the South American coast). Before 1957, when most Antarctic data first became available, there are essentially no data at all for the globe south of 45°S (refs 25, 26). Although this represents only ~15% of the area of the globe, temperature fluctuations at high latitudes are known to be larger than at lower latitudes and so can have a disproportionate effect on the global average^{12,27}. Any interpretation of Fig. 5 must bear in mind both these basic data deficiencies and the marine data uncertainties implied by

Fig. 4. We note, however, that the latter do not affect the gross features of the global mean changes observed this century.

The global curve is extremely interesting when viewed in the light of recent ideas of the causes of climatic change^{1,2}. The data show a long timescale warming trend, with the three warmest years being 1980, 1981 and 1983, and five of nine warmest years in the entire 134-yr record occurring after 1978. With regard to the hypothesized warming due to increasing concentrations of carbon dioxide and other greenhouse gases, the overall change is in the right direction and of the correct magnitude^{1,7,28}. However, the relatively steady conditions maintained between the late 1930s and mid 1970s requires either the existence of some compensating forcing factor or, possibly, a lower sensitivity to greenhouse gas changes than is generally accepted.

We thank particularly C. K. Folland, D. E. Parker, P. M. Kelly, T. P. Barnett and D. J. Shea for useful comments. We thank D. E. Parker and C. K. Folland (UK Meteorological Office) for access to unpublished data used in Fig. 4 Scott Woodruff (Environmental Research Laboratories, US National Oceanic and Atmospheric Administration) for an early copy of the trimmed COADS data set, and the Climate Analysis Center, US National Oceanic and Atmospheric Administration, for the recent marine data used in Fig. 5. This work was funded by the Carbon Dioxide Research Division, US Department of Energy, grant no. DE-FGO2-86-ER60397.

Received 10 February; accepted 28 May 1986.

1. Wigley, T. M. L., Angell, J. K. & Jones, P. D. in *Detecting the Climatic Effects of Increasing Carbon Dioxide* (eds MacCracken, M. C. & Luther, F. M.) 55–90 (U.S. Dept of Energy, 1985).
2. Hansen, J. *et al. Science* **213**, 957–966 (1981).
3. Shearman, R. J. *Met. Mag.* **112**, 1–10 (1983).
4. Folland, C. K., Parker, D. E. & Kates, F. E. *Nature* **310**, 670–673 (1984).
5. Woodruff, S. D. in *Proc. 3rd Conf. on Climate Variations and Symp. on Contemporary Climate: 1850–2100*, 14–15 (American Meteorological Society, Boston, 1985).
6. Slutz, R. J. *et al. Comprehensive Ocean–Atmosphere Data Set Release 1* (NOAA Environmental Research Laboratories, Boulder, 1985).
7. Wigley, T. M. L., Jones, P. D. & Kelly, P. M. in *The Greenhouse Effect, Climatic Change, and Ecosystems* (eds Bolin, B., Döös, B. R., Jäger, J. & Warrick, R. A.) (SCOPE series, Wiley, in the press).
8. Paltridge, G. W. & Woodruff, S. *Mon. Weath. Rev.* **109**, 2427–2434 (1981).
9. Paltridge, G. W. *Mon. Weath. Rev.* **112**, 1093–1095 (1984).
10. Barnett, T. P. in *Detecting the Climatic Effects of Increasing Carbon Dioxide* (eds MacCracken, M. C. & Luther, F. M.) 91–107 (U.S. Dept of Energy, 1985).
11. Jones, P. D. *et al. J. Clim. appl. Met.* **25**, 161–179 (1986).
12. Jones, P. D., Raper, S. C. B. & Wigley, T. M. L. *J. Clim. appl. Met.* **25** (in the press).
13. Bradley, R. S., Kelly, P. M., Jones, P. D., Diaz, H. F. & Goodess, C. A *Climatic Data Bank for the Northern Hemisphere Land Areas, 1851–1980*, DoE Tech. Rep. No. TR017 (U.S. Dept of Energy, 1985).

14. Bradley, R. S. & Jones, P. D. in *Detecting the Climatic Effects of Increasing Carbon Dioxide* (eds MacCracken, M. C. & Luther, F. M.) 29–53 (U.S. Dept of Energy, 1985).
15. Wright, P. B. *Mon. Weath. Rev.* **114** (in the press).
16. Folland, C. K., Parker, D. E. & Newman, M. in *Proc. 9th Annual Climate Diagnostics Workshop*, 70–85 (U.S. Dept. of Commerce, NOAA, Washington DC, 1984).
17. Folland, C. K. & Kates, F. E. in *Milankovitch and Climate* (eds Berger, A., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 721–727 (Reidel, Dordrecht, 1984).
18. Barnett, T. P. *Mon. Weath. Rev.* **112**, 303–312 (1984).
19. Cayan, D. R. *Mon. Weath. Rev.* **108**, 1293–1301 (1980).
20. Saur, J. F. T. *J. appl. Met.* **2**, 417–425 (1963).
21. James, R. W. & Fox, P. T. *Comparative sea-surface temperature measurements*, Marine Science Affairs Rep. No. 5, WMO Publ. No. 336 (Geneva, 1972).
22. Kirkaldy, A. W. *British Shipping, Its History, Organization and Importance* (Kegan Paul, Trench, Trubner, London, 1919).
23. Krümmel, O. *Handbuch der Ozeanographie*, Vol. 1 (Engelhorn, Stuttgart, 1907).
24. Brooks, C. F. *Mon. Weath. Rev.* **54**, 241–253 (1926).
25. Raper, S. C. B., Wigley, T. M. L., Mayes, P. R., Jones, P. D. & Salinger, M. J. *Mon. Weath. Rev.* **112**, 1341–1353 (1984).
26. Raper, S. C. B., Wigley, T. M. L., Jones, P. D., Kelly, P. M., Mayes, P. R. & Limbert, D. W. S. *Nature* **306**, 458–459 (1984).
27. Kelly, P. M., Jones, P. D., Sear, C. B., Cherry, B. S. G. & Tavakol, R. K. *Mon. Weath. Rev.* **110**, 71–83 (1982).
28. Wigley, T. M. L. & Schlesinger, M. E. *Nature* **315**, 649–652 (1985).
29. Reynolds, R. W. & Gemmill, W. H. *Trop. Ocean–Atmos. Newslett.* No. 23, 4–5 (1984).

LETTERS TO NATURE

Observation of terrestrial orbital motion using the cosmic-ray Compton–Getting effect

D. J. Cutler* & D. E. Groom*

Department of Physics, University of Utah, Salt Lake City, Utah 84112, USA

* Present addresses: Instrumentation Laboratory, North 3939 Freya, Spokane Washington 99207, USA (D.J.C.); Supercollider Central Design Group, LBL 90–4040, Berkeley, California 94720, USA (D.E.G.)

Using underground observations, we have found a small diurnal amplitude modulation of the cosmic-ray muon intensity which agrees in amplitude and phase with a first-order relativistic effect due to the Earth's motion, as discussed by Compton and Getting more than fifty years ago. The parent particles are sufficiently rigid (~1.5 TeV/c) that solar and geomagnetic effects should be minor. The muon flux deep underground is relatively insensitive

to near-surface meteorological effects, and temperature effects at production height would produce intensity variations nearly out of phase with the observed effect. Analysis of the arrival times of 5×10^8 muons during a period of 5.4 yr yields a fractional amplitude variation of $2.5^{+0.7}_{-0.6} \times 10^{-4}$, with a maximum near dawn, at 08:18 $\pm$ 1.0 h local mean solar time (LT). The expected amplitude is 3.40×10^{-4} , with the maximum at 06:00 LT.

Compton and Getting¹ showed that a cosmic-ray detector with an energy threshold would observe an enhanced intensity when it moved along its direction of maximum sensitivity with respect to the rest frame of the cosmic-ray plasma. If the cosmic-ray energy distribution were a power law of the form $E^{-\gamma}$, then the fractional intensity enhancement above a fixed energy threshold should be

$$\frac{\Delta I(\theta)}{\langle I \rangle} = (2 + \gamma) \frac{v}{c} \cos \theta$$

where θ is the angle between the direction of detector sensitivity and its velocity vector. A term v/c arises because the detector sweeps out a column of the cosmic-ray plasma, another term $2v/c$ because the solid angle transformation increases the intensity in the direction of motion, and a term $(\gamma - 1)v/c$

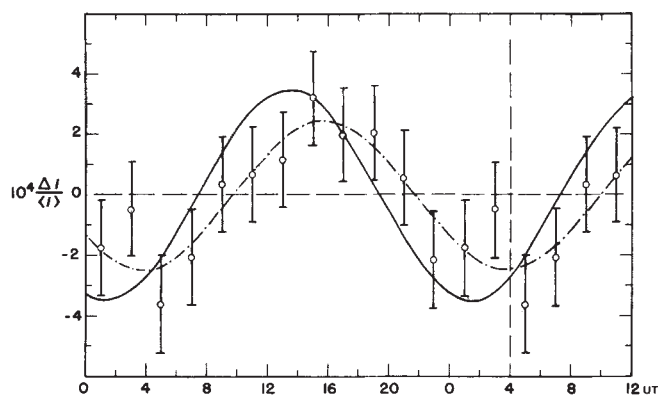


Fig. 1 Fractional muon intensity in 2-h intervals as a function of Universal Time. The dotted line is a two-parameter least-squares fit to the data; the solid curve is the *a priori* expectation due to the Earth's orbital motion. The error bars indicate a standard deviation of $\sqrt{12/N_i}$, where a total of N_i events were recorded.

because the Doppler shift of the energy spectrum changes the number of particles that exceed the detector's energy threshold. For a terrestrial detector at latitude λ sensitive to normally incident particles, and for $\gamma = 2.75 \pm 0.03$ (ref. 2), the Earth's orbital velocity should lead to a diurnal amplitude modulation of $4.75 \times 10^{-4} \cos \lambda \cos^2(i/2)$, where i is the inclination of the Earth's axis. To first order in eccentricity there are no eccentricity effects. The effect should peak near dawn (06:00 LT), when the detector is sensitive in a direction most nearly parallel to that of the Earth's velocity.

The detector, located in a mine near Heber, Utah, was designed to measure the sidereal anisotropy of the cosmic-ray intensity, and is described in a report³ of the analysis of the first 3-yr data segment. Three layers of 100 plastic scintillators each were separated by concrete absorbers. The 25×154 -cm counters were arranged in the north-south direction to optimize hour-angle resolution. A detailed readout of triple-coincidence events permitted recognition of single muons at various angles. The muon threshold energy, as intensity-averaged over the topography, was 128 GeV. This energy corresponds to a median primary rigidity (momentum per unit charge) of 1,520 GeV/c (ref. 4), if 79% of the primaries are assumed to be protons. At this rigidity, the cyclotron radius of a particle at the Earth's orbit would be ~ 6 AU, and in interstellar space ($\mu \approx 2 \mu\text{G}$) about 170 AU, or 10^{-3} pc. Because of overburden details, the effective position of the detector was displaced 1.1° north of its actual position (to 41.7° N) and 9.3° east (to 110.8° W). As a consequence, 06:00 LT occurred at 13:24 Universal Time (UT), and at this latitude the expected fractional amplitude of the Compton-Getting effect should be 3.40×10^{-4} .

Data were recorded with an average muon trigger rate of 4.7 s^{-1} between 4 January 1978 and 20 May 1983. The detector was operational 64% of the time; most of the down-time was the result of interrupted or ill-conditioned power. During this period 4.8×10^8 muons were recorded in 60,500 half-hour summary records. After correcting for any dead channels⁵ and re-binning to correct for the hour-angle differences between different angular bins, the data were Fourier-analysed to search for spectral features near harmonics of $(24 \text{ h})^{-1}$. As a check, the data were also folded with sidereal and solar periods. Such 'folded light curves' agreed well with curves synthesized from the Fourier amplitudes. Figure 1 shows the results at the solar frequency. No evidence for solar harmonics higher than the first was found. Analysis of the sidereal frequency case will be published elsewhere.

Figure 1 also shows the *a priori* expectation discussed above (solid curve) and a two-parameter least-squares fit to the data (dash-dot curve). The least-squares fit has an amplitude of

$2.47_{-0.55}^{+0.72} \times 10^{-4}$ and peaks at $15:42 \pm 1.0 \text{ h UT}$ (08:18 LT). The position-dependent amplitude is chosen from an off-centre Rayleigh distribution, as discussed in ref. 3. Quoted errors are such that the distance to the mean contains 34% of the area, in analogy to ± 1 standard deviation (s.d.) errors for a normal distribution. The χ^2 for this fit is 3.8, which is unusually small for 9 degrees of freedom (d.f.). The probability of χ^2 being this small or less is $\sim 8\%$. The data processing algorithms were tested for possible smoothing effects by processing Monte-Carlo-generated data with gaussian noise. As analysis of these data resulted in χ^2 s with expected values, we can only conclude that the low χ^2 observed with the real data is the result of a statistical fluctuation.

What other effects could confuse or mimic the Compton-Getting effect? Bending in solar magnetic fields should not be important at this rigidity, which was intentionally chosen to minimize such effects for the sidereal signal. The persistence of the sidereal signal reported in ref. 3 is evidence that this is the case. In addition, sidereal anisotropies have been reported at 500 GeV/c (refs 6, 7) and 1,000 GeV/c (ref. 8), indicating that at even lower momenta the streaming direction in interstellar space is not lost in penetrating the Solar System. Even if directions were smeared by the solar fields, all that is required for observation of the diurnal Compton-Getting effect is that the cosmic-ray plasma moves with constant velocity in the Sun's rest frame.

According to detailed integrations by D. J. Cooke (personal communication), a normally incident 1,000-GeV/c proton at the Utah detector would have been deflected 1.2° to the east by the Earth's magnetic field. An analytical calculation with a simple dipole field model yields an effect which is slightly larger, but still negligible. However, this geomagnetic field might easily wash out the Compton-Getting effect at lower energies, where sensitivity to solar activity would also be expected.

Meteorological effects are more problematical. Most cosmic-ray muons come from meson decay in the upper atmosphere, at a mean height of ~ 200 mbar. Because the competition between interaction and decay is sensitive to air density, the muon intensity is greater when the stratosphere is warmer. More specifically, the fraction of pions with energy E_π which decay rather than interact at the mean production height is $(1 + E_\pi/B_\pi)$, where the energy scale $B_\pi \approx 112 \text{ GeV}$ is proportional to temperature⁹. When kaons are included, the total temperature coefficient of $\Delta I/I$ is $2.9 \times 10^{-3} \text{ K}^{-1}$. An effect as large as that expected from the Compton-Getting effect would thus occur if the diurnal temperature modulation of the upper atmosphere were as large as 0.1°C ; in fact, temperature effects larger than this are expected^{10,11}. Similarly, a pressure coefficient of $-5.4 \times 10^{-5} \text{ mbar}^{-1}$ is expected, mostly because the increased muon range during a high-pressure period implies a higher energy threshold and therefore reduced intensity at the detector.

Meteorological effects have already received extensive attention¹²⁻¹⁴, and we conclude that correlations between temperature and pressure in radiosonde data result in incorrect experimental coefficients¹⁵; that dominant pressure effects are not particularly correlated with the diurnal period and will thus tend to average out in a least-squares sense; and that radiosonde pre-flight setting errors of as much as $\pm 1 \text{ K}$ make detailed temperature corrections impossible. Our main evidence that temperature effects are small comes from the phase of the observed effect. A temperature-related maximum should occur several hours after solar noon, nearly out of phase with the observed effect. The amplitude reduction and slight phase lag which we observe is consistent with a temperature wave with amplitude 0.07 K and a maximum at 15:00 LT.

In principle, we might divide the data to simultaneously 'observe' two directions 90° apart, and by subtraction eliminate meteorological effects. In practice, most statistical significance is lost in the process. This approach has been used successfully at lower energies (where the counting rate is higher)¹⁶, but where

geomagnetic and solar effects are important.

The orbital motion Compton-Getting effect surely exists, and generations of cosmic-ray experiments have made corrections for it, and used evidence of some signal as evidence for freedom from solar modulation. At the same time, direct observation of the effect has been curiously rare. At 500 GeV/c, Davis and co-workers^{6,7} report consistency with the effect. Gombosi *et al.*¹⁷ report it as below noise level in their air shower work at $\sim 3 \times 10^{13}$ eV/c. The strongest evidence comes from a report by the Mt Norikura group¹⁸, who report a 1.9-s.d. effect with the expected phase in their $\sim 10^{14}$ eV/c air shower observations. The present results seem to represent the best confirmation to date of this classic effect.

This work was supported by NSF grant AST80-10727. The historical involvement of H. E. Bergeson, J. F. Davis, A. Larson, J. West and countless students is gratefully acknowledged.

Received 28 March; accepted 24 April 1986.

1. Compton, A. H. & Getting, I. A. *Phys. Rev.* **47**, 817-821 (1935).
2. Ryan, M. J., Ormes, J. F. & Balasubrahmanyam, V. K. *Phys. Rev. Lett.* **28**, 985-988 (1972).
3. Cutler, D. J., Bergeson, H. E., Davis, J. F. & Groom, D. E. *Astrophys. J.* 1166-1178 (1981).
4. Gaisser, T. K. *J. geophys. Res.* **79**, 2281-2286 (1974); *Proc. 2nd int. Cosmic-Ray Symp. Jap.* **33** (1976).
5. Groom, D. E., Loh, E. C., Nelson, H. N. & Ritson, D. M. *Phys. Rev. Lett.* **50**, 573-576 (1983).
6. Davis, S. T., Elliot, H. & Thambyahpillai, T. *Proc. 16th int. Cosmic-Ray Conf., Kyoto* **4**, 210-214 (1979).
7. Davis, S. T., Elliot, H., Marsden, R. G., Thambyahpillai, T. & Dutt, J. C. *Planet. Space Sci.* **27**, 733-738 (1979).
8. Fenton, A. G. & Fenton, K. B. *Proc. 14th int. Cosmic Ray Conf., Munich* **4**, 1482-1484 (1975).
9. Barrett, P. H., Bollinger, L. M., Cocconi, G., Eisenberg, Y. & Greisen, K. *Rev. mod. Phys.* **24**, 133-178 (1952).
10. Chapman, S. & Lindzen, R. S. *Space Sci. Rev.* **10**, 3 (1969).
11. Finger, F. G. & McInturff, R. M. *J. atmos. Sci.* **25**, 1116 (1968).
12. Dorman, L. I. *Cosmic Ray Variations* (State Publishing House, Moscow, 1957).
13. Hashim, A., Speller, R. D. & Thambyahpillai, T. *J. Atmos. terr. Phys.* **34**, 1525-1535 (1972).
14. Lyons, P. R. A., Fenton, A. G. & Fenton, K. B. *Proc. 17th int. Cosmic Ray Conf., Paris* **4**, 300-303 (1981).
15. Cutler, D. J. & Groom, D. E. *Proc. 17th int. Cosmic Ray Conf., Paris* **4**, 290-293 (1981).
16. Bercovitch, M. *Proc. 17th int. Cosmic Ray Conf., Paris* **10**, 246-249 (1981).
17. Gombosi, T., Kota, J., Somogyi, A. J. & Varga, A. *Nature* **255**, 687-689 (1975).
18. Nagashima, K. *et al. Proc. 14th int. Cosmic Ray Conf., Munich* **4**, 1503-1508 (1975).

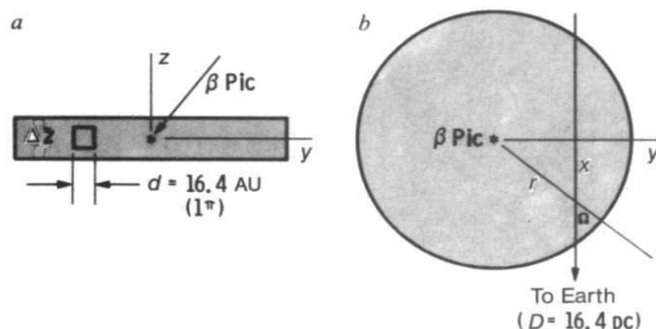


Fig. 1 Geometrical coordinates used in analysing the ground-based and IRAS observations of the disk orbiting β Pic. The disk is taken to be uniformly thick in the vertical direction and in an edge-on orientation as seen from Earth. *a*, Edge-on view; *b*, top view.

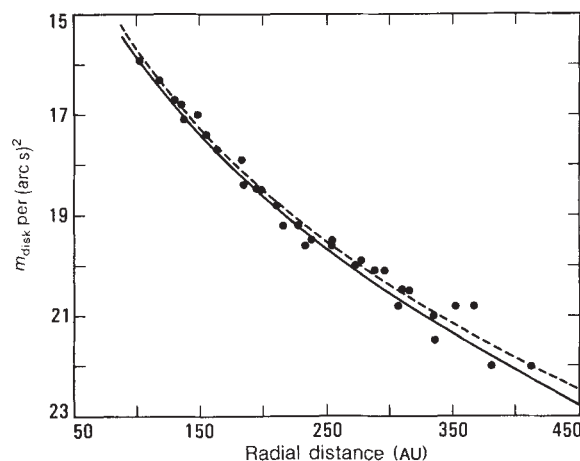


Fig. 2 Comparison of models for m_{disk} per (arc s)² as a function of radial distance from β Pic (lines) with the observations of Smith and Terile⁹ (filled circles). Dashed line, model derived from a fit to the coronagraph data only, assuming the modified gamma distribution for σ and $g=0.8$. Solid line, model derived from a simultaneous fit to the coronagraph and IRAS measurements using $g=0.8$ and $T_1=440$ K. Within the limits specified in the text, other choices for g and T_1 do not significantly affect the quality of the fit. The radial distribution parameters obtained for the models shown are $\eta=0.15$, $\gamma=0.50$ and $r_0=7.19$ AU for the 'coronagraph only' case and $\eta=2.09$, $\gamma=0.31$ and $r_0=0.03$ AU for the 'coronagraph + IRAS' case.

As a first step in modelling scattered light from the central star, it is appropriate to neglect attenuation of radiation traversing the disk and multiple-scattering effects. Thermal emission at $0.89 \mu\text{m}$ is justifiably ignored because the high radiating temperatures (for example, ~ 530 K at 200 AU) required to match the observed $0.89\text{-}\mu\text{m}$ brightness would make the disk brighter than the star at $2.2 \mu\text{m}$, a wavelength at which no flux excess is observed⁸. In addition, because molecular scattering by the gaseous envelope of hydrogen inferred spectroscopically would require an H_2 abundance greatly exceeding the upper limit of $\sim 2M_{\oplus}$ (ref. 5; $M_{\oplus}$ = mass of Earth), it is reasonable to interpret the observed features in the coronagraph data as starlight scattered by solid material. Integrating the scattering source function along the line of sight yields disk magnitude per (arc s)² as a function of distance from the star:

$$m_{\text{disk}}(y) \text{ per (arc s)}^2 = m_1 - 2.5 \log \frac{d^2}{4\pi} \int_{-\infty}^{\infty} \frac{\sigma[r(x, y)]}{x^2 + y^2} p[\Omega(x, y)] dx \quad (1)$$

Geometrical coordinates are shown in Fig. 1 and $d = 16.4$ AU

Prospecting for planets in circumstellar dust: sifting the evidence from β Pictoris

David J. Diner & John F. Appleby

Earth and Space Science Division, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, MS 183-301, Pasadena, California 91109, USA

Spectroscopic evidence for an extended gaseous shell around the A5V star β Pictoris has been available for more than a decade¹⁻⁵. Recently, the presence of a vast circumstellar cloud of particles was deduced from Infrared Astronomy Satellite (IRAS) measurements of excess thermal flux⁶⁻⁸. This interpretation is supported by a near-infrared ($0.89 \mu\text{m}$) coronagraphic image of the region beyond 100 AU from β Pic, which shows scattered starlight in a configuration consistent with a disk viewed nearly edge-on⁷. Conceivably, the β Pic system may represent one phase of planetary evolution. From an analysis of visible photometry and their coronagraph data, Smith and Terile⁹ suggested that planetary accretion has swept out an extensive (~ 30 AU) clearing at the centre of the disk. We report here a new analysis of the same data and the results of including the IRAS observations in the analysis. The consequent models do not produce a large inner clearing and suggest alternatively that significant particle number densities occur within several AU of the star. An illustrative case is used to evaluate the short-wavelength detectability of circumstellar disks in less favourable orientations than that surrounding β Pic.

(1 arc s at the distance of β Pic from Earth), σ (AU^{-1}) is the radially varying scattering coefficient, p is the single-scattering phase function, assumed dependent only on the scattering angle Ω , and m_1 is the I (0.89- μm) magnitude of the star. We use $m_V = 3.85$ for the visual magnitude of β Pic¹⁰, and $m_V - m_1 = 0.23$ for an A5V star¹¹.

Smith and Terrile's analysis⁹ of the coronagraph data relied on several assumptions: an edge-on disk, a power-law distribution for the scattering coefficient (that is, $\sigma \propto r^{-\epsilon}$) and isotropic scattering. In addition, because the absolute visual magnitude M_V of β Pic (2.78, based on a distance D of 16.4 pc) is 0.8 mag fainter than a typical A5 main-sequence star and 0.3 mag fainter than an A5 star on the zero-age main sequence (ZAMS)^{11,12}, they adopted a mean extinction of 0.5 mag relative to stars of the same type and luminosity class. This implies an optical depth of 0.46 in front of the star. Finally, because the power law is singular at the origin, they assumed $\sigma = 0$ for radial distances within an inner boundary whose location is governed by the adopted line-of-sight opacity in front of the star. From this analysis they concluded that the inner ~ 30 AU of the disk is relatively clear and that there is a fairly abrupt and substantial increase in disk opacity just outside the inner clear zone, beyond which σ decreases as the inverse cube of the radial distance. Under the assumption that the relative variation of σ is representative of changes in particle number density, a highly efficient mechanism must be invoked to explain the extensive clearing. As the inferred size of the clear region is of similar extent to our own Solar System, it is tempting to speculate that this paucity of material is the result of sweeping out by planetary accretion.

To test Smith and Terrile's model⁹ for uniqueness, we examined in detail several of their assumptions. First, we note that the inner clearing was artificially imposed *a priori* by the need to remove the singularity in the power law. To avoid this problem, we employed a continuous and bounded function for σ , namely, the modified gamma distribution¹³:

$$\sigma(r) = \sigma'(r/r_0)^\eta \exp[-(r/r_0)^\gamma] \quad (2)$$

The flexibility of this function allows for varying degrees of depletion at small r and rates of decay at large r . Second, instead of isotropic scattering, we used the Henyey-Greenstein phase function $p = (1 - g^2)(1 + g^2 - 2g \cos \Omega)^{-3/2}$ to represent more realistically the scattering by particulates¹⁴. In this expression, g is an asymmetry parameter equal to the average scattering angle cosine. Third, we examined all A5V stars in the *Bright Star Catalogue*¹⁰ with parallaxes > 0.02 arc s (that is, within 50 pc, for which interstellar extinction is negligible). We find the cumulative mean and standard deviation of M_V for stars closer than 20, 30, 40 and 50 pc to be 3.45 ± 0.52 , 2.60 ± 0.81 , 2.53 ± 0.75 and 2.27 ± 0.73 , respectively, showing that the dispersion in M_V is comparable to or larger than the deviation of β Pic from the typical main-sequence and ZAMS values. Combined with the fact that the uncertainty of ± 2.9 pc in the distance to β Pic (D. Hoffleit, personal communication) corresponds to an additional 0.4 mag uncertainty in M_V , we ascribe no statistical significance to the 0.5 mag extinction adopted by Smith and Terrile⁹. Noting furthermore that a correct assessment of the effect of disk opacity on M_V must include the scattering of light towards Earth in addition to attenuation of the direct stellar beam, we argue that the visible photometry does not provide a useful constraint on the optical properties of the disk. Instead, we included IRAS observations in our analysis.

The measured infrared magnitudes of β Pic are 2.66, 0.04, -2.75 and -3.41 at 12, 25, 60 and 100 μm , respectively⁸. The magnitude scale is defined such that 0 mag at any wavelength corresponds to the flux from a black body at $T_{\text{ref}} = 10,000$ K radiating into a solid angle $S_{\text{ref}} = 1.57 \times 10^{-16}$ ster. Using $T_* = 8,500$ K as the effective temperature of an A5V star¹¹, we calculate the solid angle of β Pic from the K (2.2- μm) magnitude⁸ of 3.47. The result implies $R_*/R_\odot = 1.17$ (where R_* and $R_\odot$

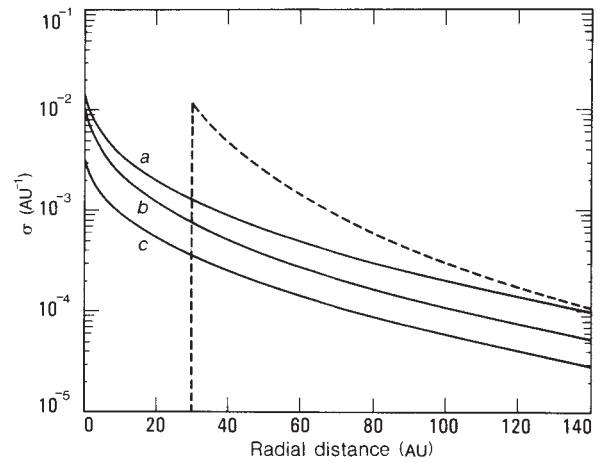


Fig. 3 Scattering coefficient σ at 0.89 μm as a function of radial distance from β Pic, using the modified gamma distribution and fitting only the coronagraph data with several choices of g (a, 0.9; b, 0.8; c, 0.6). Smith and Terrile's model⁹, also derived using only their coronagraph data, is shown for comparison (dashed line) and predicts $\sigma = 1.17 \times 10^{-2} \text{ AU}^{-1}$ at 30 AU, no material inward of this distance, and an r^{-3} dependence beyond 30 AU. (These values are taken directly from ref. 9. We note, however, that use of Smith and Terrile's model assumptions and our equation (1) implies values of σ approximately a factor of five smaller.) Given the lack of direct near-infrared measurements within 100 AU, this figure demonstrates the sensitivity of the derived radial opacity distribution to the assumed functional form used to fit the data.

are, respectively, the radius of β Pic and that of the Sun). Because the IRAS fields of view are large relative to the disk dimensions implied by the coronagraph data, the measured infrared flux is equal to the sum of the stellar flux and the contribution from the entire disk. For an optically thin disk in which each element radiates isotropically, the orientation is immaterial and the simplest method of calculating the disk flux is to imagine a face-on rather than edge-on view. Letting B_λ denote the Planck function at wavelength λ , T_d the temperature of the disk at radius r , and Δz the vertical geometric thickness, and assuming for simplicity that the infrared absorption coefficient α is wavelength-independent at 12–100 μm and follows the same radial distribution as the near-infrared scattering coefficient σ (see equation (2)), we derive model infrared magnitudes

$$m_\lambda = -2.5 \log \left\{ \frac{\exp(c_2/\lambda T_{\text{ref}}) - 1}{D^2 S_{\text{ref}}} \left[\frac{\pi R_*^2}{\exp(c_2/\lambda T_*) - 1} + 2\pi r_0 \alpha' \Delta z \int_0^\infty \frac{(r/r_0)^{\eta+1} \exp[-(r/r_0)^\gamma] dr}{\exp[c_2/\lambda T_d(r)] - 1} \right] \right\} \quad (3)$$

where $c_2 = 1.43883$ cm K. Radiative equilibrium between the dust grains in the disk and the central star implies $T_d = T_* r^{-1/2}$ (r in AU), where the temperature at 1 AU, T_1 , depends on the optical properties of the disk material.

The iterative method of nonlinear least-squares¹⁵ was used to fit, first, the coronagraph data only and, second, the coronagraph and IRAS data simultaneously. At each iteration step, the 0.89- μm solution was normalized to 19.5 mag (arc s)⁻² at 250 AU and the product $\alpha' \Delta z$ was determined by averaging the values required to match the 25- and 60- μm observations. All integrals were evaluated numerically. Individual models were characterized by specifying values for the Henyey-Greenstein asymmetry parameter and the disk temperature at 1 AU. Due to the variety of mechanisms which promote removal of very small particles over long timescales, we assume grain sizes $\geq$ visible wavelengths, in which case the limits¹⁶ $0.6 \leq g \leq 0.9$ and the range^{17,18} $200 \leq T_1 \leq 600$ K encompass a wide variety of particle

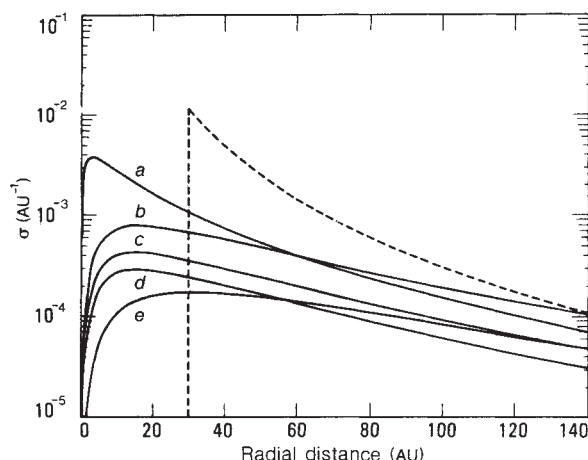


Fig. 4 Scattering coefficient σ at $0.89 \mu\text{m}$ as a function of radial distance from β Pic, using the modified gamma distribution and fitting the coronagraph and IRAS data simultaneously for various choices of g and T_1 : *a*, ($g=0.8$, $T_1=200$ K); *b*, (0.9 , 440); *c*, (0.8 , 440); *d*, (0.6 , 440); *e*, (0.8 , 600). The distribution at small radial distances is governed primarily by the IRAS results; note that a cooler disk requires larger opacity near the star to fit those data. Increased disk opacity is also required for higher values of g , to compensate for the decrease in amount of starlight scattered towards Earth. As an example of the magnitude of derived disk opacities, we use the case with $g=0.8$ and $T_1=440$ K (model *c*) for illustration. We calculate a peak vertical (z -direction) infrared absorption optical depth of 1.5×10^{-3} , which occurs at 15 AU from the star in this model. The corresponding scattering opacity in the near-infrared depends on an assumed value for Δz : we compute in this case a peak vertical near-infrared scattering optical depth of 4.3×10^{-3} for each 10 AU of disk thickness. Note that in models *a*–*e*, significant disk opacity (50% of the peak value) is found within several AU of the star. Smith and Terrile's model⁹ is shown for comparison (dashed line).

sizes and complex refractive indices. The range for T_1 has been scaled from calculations performed for studies of the zodiacal cloud. Note that the value $T_1=440$ K is the equilibrium temperature for black-body grains.

The least-squares iteration was ended following negligible improvement in the sum of the squares of the residuals between the measurements and corresponding model magnitudes. Comparisons of our model $0.89\text{-}\mu\text{m}$ magnitudes with the coronagraph data indicate excellent fits (see Fig. 2). When the IRAS data are excluded from the fitting procedure, the predicted infrared magnitudes, not surprisingly, do not match the IRAS data. For example, the model with $g=0.8$ yields 12-, 25-, 60- and $100\text{-}\mu\text{m}$ magnitudes of 0.80 , -0.78 , -2.28 and -2.89 , respectively. Choosing $T_1=440$ K for illustration and including the IRAS data, these values become 2.55 , -0.05 , -2.66 and -3.66 mag. This is a satisfactory fit, especially in view of our simple assumptions. Thus, the introduction of further complexity into the modelling is not warranted until additional observations (such as spatially

resolved infrared data or coronagraphic images with greater radial coverage) become available. We note that the goodness-of-fit to the IRAS data is not particularly sensitive to g and that the warmer-disk models give better fits at $12\text{--}100 \mu\text{m}$. The models which incorporate the IRAS data predict no flux excess at $2.2 \mu\text{m}$ and ~ 0.02 mag at $5 \mu\text{m}$.

The scattering coefficient at $0.89 \mu\text{m}$ derived by fitting the coronagraph data only is shown in Fig. 3 for several values of g . Figure 4 shows $\sigma(r)$ derived from simultaneous fits of the coronagraph and IRAS data for several combinations of g and T_1 . The Smith and Terrile model⁹ is reproduced in both figures. Note that for all of our models, significant opacity occurs well within the region which they inferred to be swept clear of material. Thus, these models indicate no sizeable inner clearing, particularly of the scale inferred by Smith and Terrile. It remains an open question whether planetary formation must be invoked to explain the distribution of disk material. Other mechanisms could account for a decrease in opacity close to the star, such as loss of particles due to radiative, electromagnetic or aerodynamic effects; or a change in particle properties such as a decrease in mean size.

One practical application of combined modelling of scattered and emitted radiation is in predicting the detectability of randomly oriented circumstellar disks by coronagraphic imaging. The nearly edge-on orientation of the β Pic disk provides the most favourable condition for detection of scattered starlight. Taking m_{disk} per (arc s)² < 22 as a reasonable criterion for detection (see Fig. 2) and using the case with $g=0.8$ and $T_1=440$ K for illustration, we calculate that if the β Pic disk were in the less favourable face-on orientation, it would be detectable by coronagraphic imaging at 100 AU from the star if its vertical geometric thickness (Δz) is greater than a few tenths of an AU and at 200 AU radial distance if $\Delta z \geq 4$ AU. IRAS has identified a number of stars with thermal infrared excesses, but with the exception of β Pic, attempts to obtain optical images of the material around those stars have, to our knowledge, not been successful.

In future, simultaneous acquisition of scattered-light and thermal infrared observations of stars with circumstellar shells or disks should be a promising method of determining the nature and distribution of orbiting material. Spatially resolved infrared observations, perhaps obtained interferometrically, combined with multi-spectral coronagraphic images with smaller occulting masks, would be very useful in characterizing circumstellar disk structure. The degrading effects of the terrestrial atmosphere probably dictate acquisition of such data from a spaceborne platform. Determination of the ubiquity of objects such as β Pic, evaluation of whether orientation of circumstellar disks is a critical factor governing their detectability, and placement of such systems into the context of planetary evolution must await the collection and analysis of further infrared and coronagraphic observations.

We thank H. H. Aumann, J. Diner, M. S. Hanner, J. V. Martonchik, D. J. McCleese and J. A. Russell for helpful discussions and W. Wells for care in preparation of the manuscript. This work was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

Received 13 January; accepted 7 April 1986.

1. Sletteback, A. *Astrophys. J.* **197**, 137–138 (1975).
2. Sletteback, A. *Astrophys. J. Suppl. Ser.* **50**, 55–84 (1982).
3. Sletteback, A. & Carpenter, K. G. *Astrophys. J. Suppl. Ser.* **53**, 869–892 (1983).
4. Kondo, Y. & Bruhweiler, F. C. *Astrophys. J.* **291**, L1–L5 (1985).
5. Hobbs, L. M., Vidal-Madjar, A., Ferlet, R., Albert, C. E. & Gry, C. *Astrophys. J.* **293**, L29–L33 (1985).
6. Aumann, H. H. *Bull. Am. astr. Soc.* **16**, 483 (1984).
7. Aumann, H. H. *et al. Astrophys. J.* **278**, L23–L27 (1984).
8. Aumann, H. H. *Publ. astr. Soc. Pacif.* **97**, 885–891 (1985).
9. Smith, B. A. & Terrile, R. J. *Science* **226**, 1421–1424 (1984).

10. Hoffleit, D. & Jaschek, C. *The Bright Star Catalogue* (Yale University Observatory, New Haven, 1982).
11. Allen, C. W. *Astrophysical Quantities* (Athlone, London, 1981).
12. Crawford, D. L. *Astr. J.* **84**, 1858–1865 (1979); **85**, 621–622 (1980).
13. Deirmendjian, D. *Electromagnetic Scattering on Spherical Polydispersions* (Elsevier, New York, 1962).
14. Irvine, W. M. *Icarus* **25**, 175–204 (1975).
15. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969).
16. Hansen, J. E. & Travis, L. D. *Space Sci. Rev.* **16**, 527–610 (1974).
17. Lamy, Ph. L. *Astr. Astrophys.* **35**, 197–207 (1974).
18. Röser, S. & Staude, H. J. *Astr. Astrophys.* **67**, 381–394 (1978).

Radio observations of PKS2314+03 during occultation by comet Halley

S. K. Alurkar, R. V. Bhonsle & A. K. Sharma

Physical Research Laboratory, Ahmedabad-380009, India

Strong interplanetary scintillations (IPS) were recorded at 103 MHz at Thaltej (23°02' N, 72°36' E) between 18 and 20 December 1985, when the quasar PKS2314+03 was occulted by the ion tail of comet Halley. The scintillation periodicities were near 1 s and their amplitudes decreased progressively as the source approached the end of the tail. The scintillating flux density on these days was 18.2, 11.4 and 4.7 Jy when the source was ~86° from the Sun. The r.m.s. density variations were 10, 6 and 3 electrons cm⁻³ on 18, 19 and 20 December respectively, and varied with distance r from cometary nucleus as $r^{-3.3}$. Assuming an ion velocity of 100 km s⁻¹, we show here that the observed periodicity of 1 s implies ion density inhomogeneities of the order of 100 km.

Comets develop ion or type I tails when they are 2–4 AU from the Sun¹. Plasma features such as kinks and knots seen in these tails are caused through interaction with the solar wind^{2,3}. Plasma density inhomogeneities in the solar wind scatter radio waves from a compact radio source and produce interplanetary scintillations that apparently broaden the radio source⁴. Attempts have been made to observe the angular broadening of radio sources as they were occulted by comet Arend-Roland⁵, and scintillations were observed during the occultation of radio source PKS2025–15 by comet Kohoutek⁶.

Regular observations of seven radio sources in the declination range +2° to +23° were made during December 1985 at Thaltej using a 10,000-m² correlation interferometer operating as a transit instrument at 103 MHz (ref. 7). Of these, six were scintillating sources. Sine and cosine outputs of the correlation receiver, together with a scintillometer⁸ output, were recorded

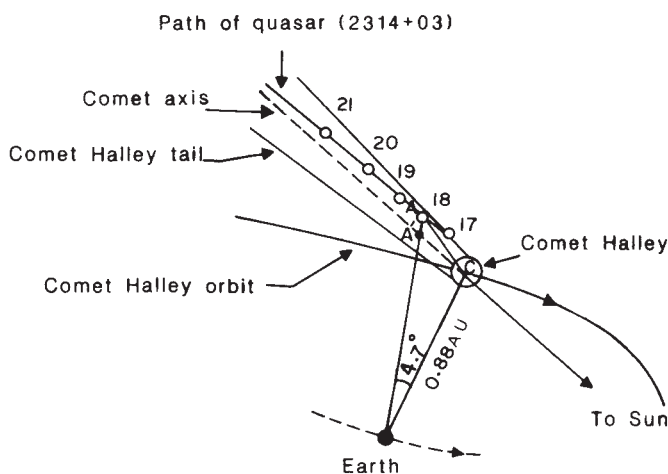


Fig. 1 Geometry, not to scale, of the occultation of PKS2314+03 (3C459) by comet Halley.

on a strip chart. These outputs, sampled 20 times per second, were recorded with time information on a digital magnetic tape. A real-time fast Fourier transform spectrum analyser (Model 512/S Rockland) can provide an average power spectrum of the scintillating source over a frequency range of 0.1–20 Hz.

Comet Halley was predicted to occult PKS2314+03 (or 3C459) during the period 17–21 December 1985. 3C459 is a scintillating source with ~70% of the total flux in the 0.45 arc s component⁹.

Figure 1 shows the geometry of the occultation for the period 17–21 December, for which the path of 3C459 as seen through the comet tail is indicated. The comet's position angle on 18 December was ~66°, and the quasar on that day was 87.5° away from the Sun. Some relevant parameters of the occultation are summarized in Table 1. 3C459 has been observed since the

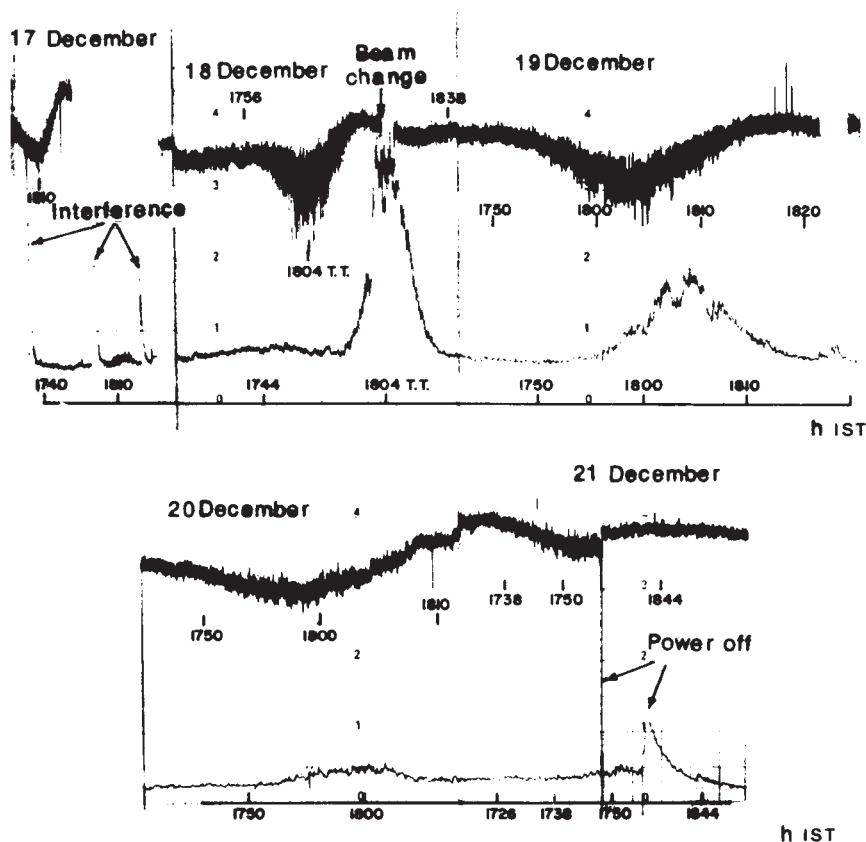


Fig. 2 103-MHz recordings of 3C459 during 17–21 December 1985. Top trace in each panel, sine receiver output; bottom trace, scintillometer output. Chart speed on 17 and 18 December was 5 and 10 cm h⁻¹, respectively, and 20 cm h⁻¹ thereafter. Note the strong scintillations on 18 and 19 December and weaker scintillations on 20 December.

beginning of December 1985, when it was $\sim 103^\circ$ from the Sun. Its average scintillating flux density was 3.5 Jy until 17 December, when the source was observed through the cometary tail at a distance of 0.08 AU from its nucleus.

On 18 December, however, as shown in Fig. 2, very intense scintillations were recorded when the distance from the comet nucleus to the point of intersection of the line of sight with the tail (AC in Fig. 1) was 0.12 AU. On 19 and 20 December the scintillations decreased progressively, when the corresponding distances were ~ 0.14 and 0.18 AU. Note that only sine and scintillometer outputs are shown in the recordings of Fig. 2, as the deflection due to 3C459 appears only on the sine response of the correlation receiver⁷. On 21 December, when the source-comet distance was 0.2 AU, the scintillating flux density of 3C459 decreased to ~ 3.3 Jy.

Following the calibration procedure suggested by Duffett-Smith¹⁰ for estimating the mean source flux and the scintillating flux, IPS observations on 18–21 December, shown in Fig. 2, were analysed. The scintillometer outputs on these days correspond to r.m.s. scintillating flux densities of 18.2 ± 0.04 , 11.4 ± 0.04 , 4.7 ± 0.04 and 3.3 ± 0.04 Jy. Thus, the scintillation index, m (= r.m.s. scintillating flux/mean source flux) values become 0.42, 0.27, 0.11 and 0.08 respectively, for a mean source flux of 43 Jy.

Note that the average scintillating flux of 3C459 when it is $\sim 90^\circ$ away from the Sun is usually 3.5 Jy and reaches a maximum of ~ 12 Jy at $28\text{--}30^\circ$, when m begins to turn over at 103 MHz. Allowing for day-to-day maximum variations of 2–3 times 3.5 Jy, the measured scintillating flux values during the occultation by comet Halley are very significant.

The scintillation spectra (computed by the real-time spectrum analyser) obtained between 17 and 20 December are shown in Fig. 3. These were normalized with respect to the maximum value and are source spectra computed as the difference between 'on-source' and 'off-source' spectra, each averaged over 10 min. The spectra resemble typical IPS spectra¹¹, with a nearly flat low-frequency region below the spectral variations at ~ 0.2 Hz on 17 and 20 December, and at ~ 0.6 and 0.4 Hz on 18 and 19 December, respectively, when the scintillations were very strong. Auto-correlation analysis of the spectra gave average scintillation periodicities of 1.0 s.

The solar and ionospheric disturbances which might have caused these scintillations were studied using the data for the period 2–22 December 1985 (see ref. 12). Solar activity throughout this period was 'very low', while geomagnetic activity was 'quiet to unsettled', with no Sc-type magnetic storms. Furthermore, ionospheric beacon satellite data at very high frequency and ionograms at Ahmedabad did not indicate scintillation activity and spread-F respectively during this period.

An interplanetary transient or ionospheric scintillation causing the enhancement in scintillation of 3C459 was ruled out by studying IPS of 3C298, 3C318, 3C324, 3C368, 3C409 and 3C459, covering the declination range $+2^\circ$ to $+23^\circ$, which were being regularly observed. Of these, 3C298 and 3C368 are within 8° declination of 3C459. Using many scintillating sources, Hewish *et al.*¹³ derived the shape and dynamics of large-scale interplanetary transients, which cover a typical solid angle of helio-

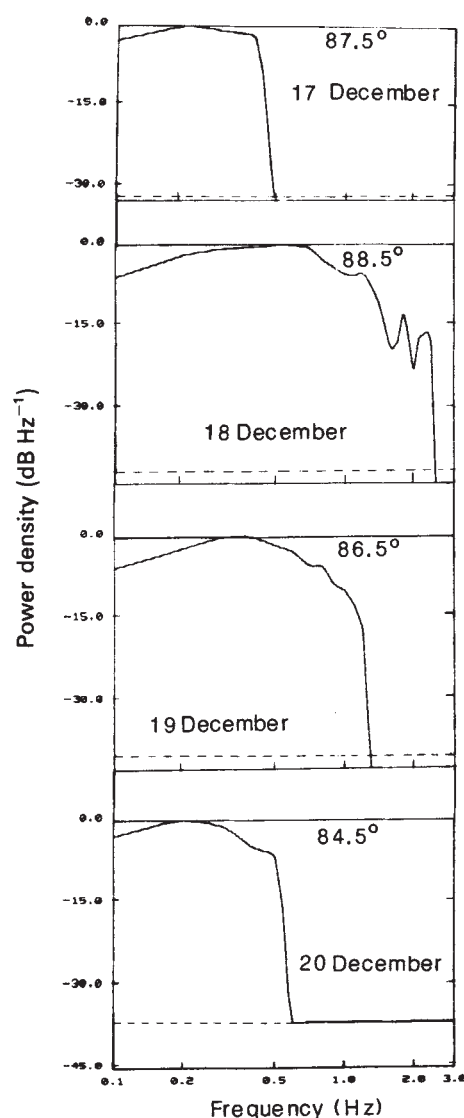


Fig. 3 Scintillation spectra of 3C459 on 17–20 December 1985. Dashed lines indicate noise level. Numbers in upper right-hand corners are solar elongations of 3C459.

longitude and latitude of $\pi/2$ sr. They also showed a strong correlation between enhancement in scintillation and total plasma density along the line of sight to a source. Except for 3C459, no other source showed enhanced scintillations during 18–20 December 1985, indicating that the enhancement was restricted to the direction of 3C459.

Thus, the strength of the observed scintillations in the direction of 3C459, and their progressive decrease from 18 to 21 December 1985 during the expected occultation of the source by comet Halley, strongly support our conclusion that the scintillations were caused by the scattering of the radio waves by density inhomogeneities in the ion tail of comet Halley.

The values of m on 18–21 December depend on the source-comet distance along the axis of the comet (A'C in Fig. 1), such that $m \propto r^{-3.3}$, where r in AU is the distance between the comet nucleus and A', the projection of A on the comet axis. For gaussian irregularities with weak scattering, and for a layer of thickness L containing density irregularities of scale size a , m is given by¹⁴

$$m^2 = 2(\pi)^{1/2} r_e^2 \lambda^2 \langle \Delta N^2 \rangle a L$$

λ is the wavelength of radiation passing through the scattering layer and r_e is the classical electron radius. In the case of comet

Table 1 Parameters of the occultation geometry

Date (December 1985)	Comet nucleus- source distance along axis (AU)	Comet- Earth- source angle (deg.)	Geocentric distance of comet (AU)	Solar elongation of 2314+03 (deg.)
17	0.08	3.4	0.86	87.5
18	0.12	4.7	0.88	86.5
19	0.14	5.9	0.90	85.5
20	0.18	7.0	0.92	84.5
21	0.20	8.1	0.94	83.5

Halley, L is the thickness of its cylindrical ion tail, which is $\sim 10^5$ km (ref. 15).

The scintillation periodicities of 1 s imply scale sizes of density inhomogeneities in the cometary ion tail of 100 km (assuming their average velocity of 100 km s^{-1} (ref. 16)), which compares well with scale sizes in the solar plasma estimated from IPS measurements¹⁷.

Using the above formula, the average values of the r.m.s. density variations, $\langle \Delta N^2 \rangle^{1/2}$, in the ion tail are 10, 6 and 3 electrons cm^{-3} on 18, 19 and 20 December 1985, respectively. If the mean ion density in the tail of comet Halley is assumed to be 10^2 ions cm^{-3} (ref. 18) then $\Delta N = 10$ corresponds to 10% modulation of the mean density. This modulation is much higher than 4% for the solar wind at 0.1 AU.

Note that on 17 December 3C459 had just begun to be occulted by the comet Halley tail at a distance of ~ 0.08 AU from its nucleus. Average scintillations were observed on that day. This may be due to insufficient turbulence in that tail region. Also, ion density condensations in cometary tails cannot exist within 0.1 AU of the nucleus, due to the stabilizing effect of solar wind turbulence on the Kelvin-Helmholtz instabilities^{16,19}.

The length of the plasma tail of comet Halley over which scintillations were observed up to 20 December was ~ 0.10 AU or 1.5×10^7 km, which is in good agreement with optical measurements¹⁵.

We thank Professor S. P. Pandya for interest and the staff of Radio Astronomy and Ionospheric Groups for assistance. Discussions with Professors B. Buti and R. K. Varma were useful. We appreciate valuable suggestions by Dr P. J. Duffett-Smith of Cavendish Laboratory. Financial support for this work came from Departments of Space, Science & Technology, Government of India.

Received 30 January, accepted 25 April 1986.

1. Brandt, J. C. *Introduction to Solar Wind* (Freeman, San Francisco, 1970).
2. Jockers, K. & Lust, R. *Astr. Astrophys.* **26**, 113-121 (1973).
3. Ip, W.-H. & Axford, W. I. *Comets* (ed. Wilkening, L. L.) 588-636 (University of Arizona Press, 1982).
4. Hewish, A. & Symonds, M. D. *Planet. Space Sci.* **17**, 313-320 (1969).
5. Whitfield, G. R. & Hoggom, J. *Nature* **180**, 602 (1957).
6. Ananthakrishnan, S., Bhandari, S. M. & Pramesh Rao, A. *Astrophys. Space Sci.* **37**, 275-283 (1975).
7. Alurkar, S. K. et al. *J. Inst. Electron. Telecommun. Engrs* **28**, 577-582 (1982).
8. Duffett-Smith, P. J., *Mon. Not. R. astr. Soc.* **177**, 349-355 (1976).
9. Readhead, A. C. S. & Hewish, A. *Mem. R. astr. Soc.* **78**, 1-49 (1974).
10. Duffett-Smith, P. J., thesis, Cambridge Univ. (1976).
11. Scott, S. L. thesis, Univ. California, San Diego (1978).
12. Preliminary Report and Forecast of Solar Geophysical Data (Joint NOAA-USAF Space Environment Services Center, 1985).
13. Hewish, A., Tappin, S. J. & Gapper, G. R. *Nature* **314**, 137-140 (1985).
14. Little, L. T. *Meth. Exptl Phys.* **12c**, 118-137 (1976).
15. Jockers, K. *Proc. ESO Workshop, Paris* (1982).
16. Ershkovich, A. I. *Space Sci. Rev.* **25**, 3-34 (1980).
17. Readhead, A. C. S., Kemp, M. C. & Hewish, A. *Mon. Not. R. astr. Soc.* **185**, 207-225 (1978).
18. Schmidt, H. U. & Wegmann, R. *Comets* (ed. Wilkening, L. L.) 538-560 (The University of Arizona Press, Tucson, 1982).
19. Buti, B. *Astrophys. J.* **252**, L43-L47 (1982).

Imaging the outflow of ionospheric ions into the magnetosphere

Y. T. Chiu*, R. M. Robinson*, G. R. Swenson*, S. Chakrabarti† & D. S. Evans‡

* Space Sciences Laboratory, Lockheed Palo Alto Research Laboratories, Palo Alto, California 94304, USA

† Space Sciences Laboratory, University of California, Berkeley, California 94720, USA

‡ Space Environment Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado 80303, USA

The discovery of ionospheric ion outflow into the magnetosphere has been considered a major advance in magnetospheric physics in the past decade. Satellite measurements *in situ* are, however, too localized to address the global scope of the phenomenon. It is therefore important to consider methods to image the global outflow of ionospheric ions into the magnetosphere. We consider below the concept and verification of imaging the global ionospheric outflow by using spectrophotometric observations of far-ultraviolet and extreme-ultraviolet solar lines resonantly scattered by ionospheric ions or exospheric neutrals in the auroral regions of the magnetosphere. We shall also present new satellite observations which verify the concept and show that this imaging capability is feasible with present technology.

Our concept of magnetospheric imagery proposed here originated with the interpretation of high-sensitivity, high-resolution observations of the solar 304-Å emission line intensity resonantly scattered by He^+ in the plasmasphere. The observations were obtained by an extreme-ultraviolet telescope on the Apollo-Soyuz mission¹. The line intensities, measured while the instrument line of sight was spinning with the spacecraft, were shown to be in excellent agreement with line-of-sight He^+ density distributions predicted by a kinetic equilibrium model of the plasmasphere², when ionospheric densities measured simultaneously by the AE-C satellite at 300 km altitude were used as boundary conditions of the model. The success of this experiment and interpretation suggests that an image in line emissions resonantly scattered by magnetospheric ions of ionospheric origin or exospheric neutrals may similarly be converted into

an image of their density distribution in the magnetosphere. Furthermore, if spectrometric information is available, the flow speed and temperature of the ions or neutrals can be derived, so that an image of the ionospheric ion outflow into the magnetosphere in the high-latitude region can be made. (For reviews of various aspects of the ion outflow phenomenon, see refs 3-6.)

The generalization of concepts stemming from the earlier He II 304-Å work¹ requires a search for other emission lines, because the high-latitude ion outflows consist primarily of O^+ and H^+ , as well as some He^+ in the polar cap. We have determined that the solar O II 834-Å line can be used analogously to the He II 304-Å line to determine O^+ distributions and outflow. The proton has no resonant scattering lines; however, outflowing hot H^+ can charge-exchange with cold exospheric neutral atoms to produce hot outflowing H^0 which can, in turn, resonantly scatter Doppler-shifted Ly- α from the Sun. Thus, the same principle can be applied to Doppler-shifted Ly- α observations to determine the hot H^0 distribution. We show that our concept is workable, in principle, and is verifiable with current observations, although there is no imaging capability in currently available instruments. Details of data deconvolution will not be considered.

As both the He^+ and O^+ lines can be treated analogously, we discuss these two together. Figure 1 summarizes the results of observation and interpretation of the He^+ line in the plasmasphere.

For the He^+ and O^+ lines in the auroral region, there are two important sources of illumination: (1) solar line emission, F_s , and (2) auroral line emission, F_a , excited at ~ 100 km altitude by electron impact. The solar flux F_s is uniform but, as the auroral source is close to the region of resonant scattering, the geometric properties of F_a need to be included. The resonantly scattered O II 834-Å and He II 304-Å line intensities, I , seen above ~ 600 km altitude can be considered optically thin⁷, especially as the auroral ionosphere is particularly tenuous. Thus, the intensity is given by

$$I = \int dL (F_s + F_a) \sigma_{rs} N_i(L) \quad (1)$$

where σ_{rs} is the resonance scattering cross-section ($\approx 2.6 \times 10^{-13} \text{ cm}^2$ for the 834 Å line) and $N_i(L)$ is the ion density along the line of sight, L . This equation allows us to derive the

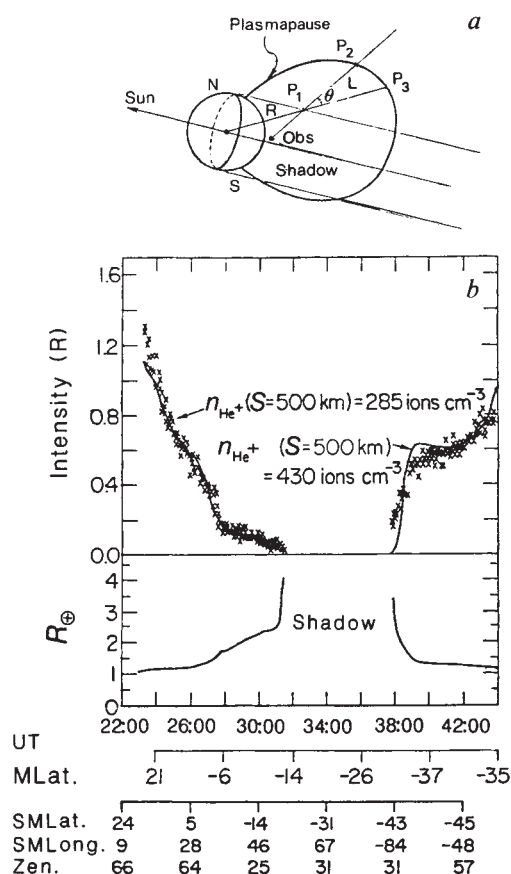


Fig. 1 Intensity of the He^+ 304-Å line measured in a spin-scan of the Apollo-Soyuz photometer. The viewing geometry is shown in *a*. The satellite (OBS) measures the line intensity from the sunlit portion of the line of sight (L) between P_1 and P_2 . The altitude (R_1) of the point P_1 which is the intercept of L with the terminator, is shown in *b*. The magnetic latitude of P_1 (MLAT), the satellite magnetic coordinates (SMLat., SMLong.) and the zenith angle of the line of sight (Zen.) are also shown. The agreement between the model (solid lines) and the data (crosses) demonstrates the capability of deriving the density distribution of He^+ in the plasmasphere by imaging the 304-Å line intensity.

appropriate ion density distribution with the help of a theoretical profile superimposed on the viewing geometry shown in Fig. 1. The peculiarities of the Earth's shadowing impose tomographic signatures (note the kinks in the line intensity profile) in much the same way as lead is used in X-ray tomography in the laboratory. A unique set of density model parameters can be determined.

We have been able to test this concept with the O II 834-Å line by using data from the University of California Berkeley extreme-ultraviolet spectrometer flown on board the P78-1 satellite⁸. The P78-1 satellite was placed in a 600-km-altitude circular orbit, with an inclination of 97.7° and an orbital period of 96 min. The orbit is Sun-synchronous, precessing at 1° per day, with the spacecraft orbit lying essentially in the noon-midnight plane. The spectrometer is housed in the spinning wheel of the spacecraft, with the spin axis of the wheel perpendicular to the orbital plane. The spectrometer's line of sight is oriented 120° from the spin axis, and, as the wheel rotates at 11 r.p.m., the instrument line of sight sweeps out a cone, alternately viewing Earth and space, never looking closer than 30° to the Sun.

To test the magnetospheric O^+ imaging model, we have chosen a pass where a night-time aurora was observed by the instrument over the South Pole on 22 March 1979. The down-looking intensity profile of the O^+ 834-Å feature as a function of latitude indicates an auroral arc at ~45,000 s UT, centred between -50°

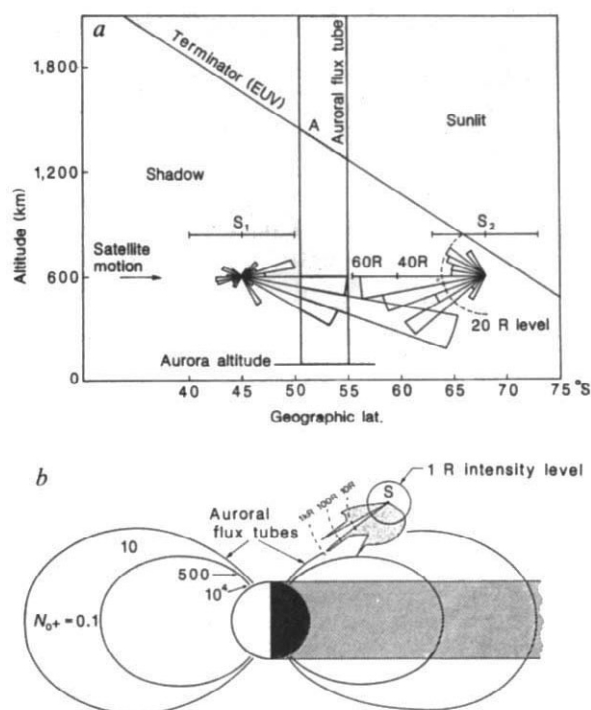


Fig. 2 *a*, Intensities of the O^+ 834-Å line as a function of the angle between the zenith and the line-of-sight direction of the extreme-ultraviolet (EUV) spectrometer aboard the satellite P78-1. Spin-angle intensity profiles accumulated during the intervals centred at S_1 and S_2 are shown. Wavelength-independent background has been subtracted. The location and extent of the auroral flux tube are determined by the measured zenith and nadir intensities. Note that the 834-Å source volumes are at altitudes as high as 2,000 km. *b*, Simulated O^+ 834-Å line intensity profile seen by a hypothetical satellite S at $4R_E$ is shown on the nightside lobe of an auroral flux tube. The input O^+ density profile⁹ as a function of distance along the field line is shown on the dayside lobe of the flux tube. Instrument sensitivity of 1 R is needed to image the distribution of O^+ on the entire auroral flux tube.

and -55° geographic latitude. Figure 2*a* shows spin-scan profiles of the O^+ 834-Å line accumulated for ~150 s, centred at two satellite locations S_1 and S_2 along the satellite path towards the South Pole, plotted as linear polar bar graphs according to the directions of the zenith-angle bins. The satellite was in the Earth's shadow at both points, viewing the auroral region forward from S_1 and then backward from S_2 . From the forward-backward anisotropy of the intensity pattern, it is clear that the instrument is viewing emissions from the auroral flux tube and that this is the predominant emission source at such high altitudes. The up-looking intensities of 1-20 R are of particular interest because they correspond to sources located at altitudes as high as 2,000 km. Using the model parameters⁹ of $N(\text{O}^+) \approx 500 \text{ cm}^{-3}$ at 2,000 km over a typical evening discrete-auroral-arc region of width ~200 km lit by solar and auroral 834-Å emissions, we found, for example, that the model predicts a luminosity of 6 R, in agreement with the up-looking intensities at S_2 , which include the sunlit and aurora-lit regions of the flux tube at 2,000 km. The extreme-ultraviolet background has been corrected for in the data and the 834-Å signal, though weak, stands out well above background.

The data clearly show that the auroral flux tube up to ~2,000 km can be imaged with an extreme-ultraviolet instrument at 834 Å. The intensity observed is in agreement with that expected from auroral flux tube models⁹. This verification allows

us to simulate what we may expect from the O^+ line when viewed from a high-altitude platform. We have run a model of electric field and plasma distributions⁹ for a discrete arc with a total potential drop of 1 keV and O^+ density of 500 cm^{-3} and H^+ density of 70 cm^{-3} at 2,000 km, to simulate an average auroral flux tube. These densities are consistent with ISIS-2 observations¹⁰. The O^+ densities are shown at roughly appropriate locations on the sunlit auroral lobe in Fig. 2b. The solar flux is assumed to be 1.3 kR, and the auroral flux is $\sim 1 \text{ kR}$ (ref. 11) at $\sim 100 \text{ km}$ altitude. With such a model, one can simulate the expected intensity of the $834\text{-}\text{\AA}$ resonance scattering line seen at any viewing geometry. Such a simulated intensity plot for the $834\text{-}\text{\AA}$ line is shown on the nightside lobe of Fig. 2b, in which the intensity of a 360° spin-scan is shown as the mottled area centred around a hypothetical satellite, S, at $4 R_E$ (where R_E is the radius of the Earth). It is seen that instrument sensitivity to 1R is required to obtain a deconvolution of O^+ density distribution of the auroral flux tube to $\sim 5 R_E$. The simulation is for the minimum outflow of an inverted V. Other sources of O^+ outflow, such as conics and polar wind, add to the $834\text{-}\text{\AA}$ intensity. Note that the simulation applies only to $834\text{-}\text{\AA}$ emission of O^+ accelerated upward by the inverted V. Such emissions have distinct Doppler shifts and broadenings. When the instrument is looking in the downward direction, the line core will be cluttered with $834\text{-}\text{\AA}$ dayglow¹² ($\sim 0.5 \text{ kR}$) and nightglow¹³ ($\sim 1 \text{ R}$). These issues of background require high-resolution spectrometry, but we cannot address these issues here.

If our concept of magnetospheric imagery is valid, it would imply that solar Ly- α emissions should also be resonantly scattered by the hot H^0 outflow created by charge-exchange interaction between the H^+ outflow and neutral constituents of the exosphere. Such emissions are clearly distinguishable from the intense geocoronal background because such hot H^0 will resonantly scatter solar Ly- α at Doppler-shifted wavelengths from the cold geocoronal line core. Note that the geocoronal Ly- α width corresponds to an exospheric temperature of $\sim 1,000 \text{ K}$ ($\sim 0.1 \text{ eV}$) for the cold H^0 ; therefore, any H^0 with higher energy is defined as hot for the present purposes. Our concept predicts the existence of a hot Ly- α emission source at the auroral magnetospheric flux tubes.

Evidence of such a hot emission source in the magnetosphere was obtained by the spectrophotometer flown on Spacelab 1. The instrument was equipped with an absorber specifically to eliminate the geocoronal Ly- α line core so as to study hot emission sources¹⁴ which are Doppler-shifted and/or broadened to the geocoronal line wings. An unexplained magnetospheric source of hot emissions with an intensity of $\sim 200 \text{ R}$, after subtraction of interplanetary background, and $\sim 20^\circ$ wide in orbital angle was detected by the instrument¹⁴ at $\sim 22:54 \text{ UT}$ on day 341 of 1983. The viewing geometry projected onto the horizontal plane is illustrated in Fig. 3, showing the Spacelab 1 ground track at southern high latitudes. The hot emission source was observed in the zenith direction, perpendicular to the ground track and out of the plane of the figure. Within a few minutes of the event, the NOAA-7 satellite flew through the southern auroral region and the NOAA-8 satellite flew through the northern auroral region. Both detected auroral electrons in the vicinity of -65° to -70° geomagnetic latitudes, but virtually no protons above the instrument threshold were found. Assuming conjugacy, these two NOAA satellite observations approximately defined the low-altitude aurora as shown on Fig. 3. The hot emission source is clearly not from the direction of the low altitude aurora, whose intensity in Ly- α is usually $\sim 6 \text{ kR}$ and is emitted by the cold geocoronal H^0 . We conjecture that the hot emission of $\sim 200 \text{ R}$ originates from Doppler-shifted Ly- α resonance scattering of solar emissions by the hot H^0 which are produced by charge exchange between outflowing hot H^+ and cold exospheric neutral atoms at several thousand kilometres up the auroral flux tube. If so, the source region is bounded below by the intercept of the equatorward edge of the auroral

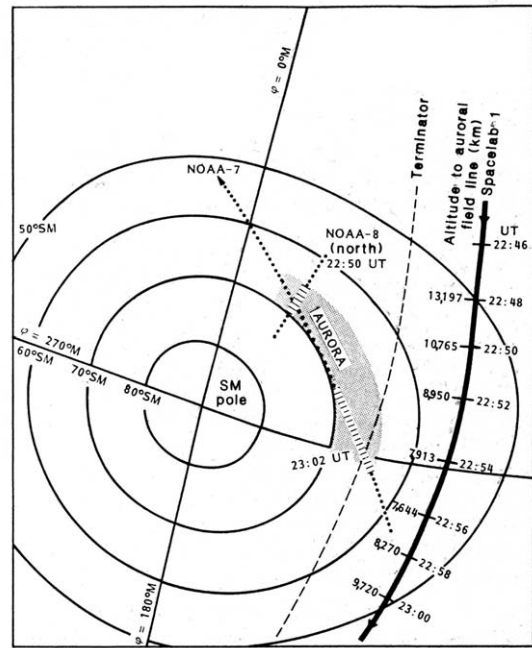


Fig. 3 Viewing geometry of Spacelab 1 observation¹⁴ of an unaccounted-for Ly- α emission source in the magnetosphere. The source was observed between 22:50 and 22:57 UT, near the zenith direction.

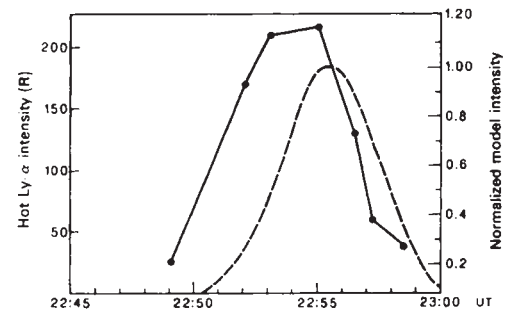


Fig. 4 Comparison of the observed hot Ly- α intensity profile (solid curve) with the expected model profile (dashed curve). The model curve is normalized to unity at 22:56 UT, when Spacelab 1 is at closest approach to the equatorward boundary of the auroral flux tube in the zenith direction. The model assumes that the unknown hot Ly- α source is the hot H^0 produced by charge exchange between hot outflowing H^+ and exospheric neutrals.

flux tube with the line of sight of the instrument. The altitudes of these equatorward intercepts are shown along the Spacelab 1 ground track on Fig. 3. Note that the closest approach of these equatorward intercepts to the instrument occurs near the observed hot emission source. This intercept is important in understanding why the source is highly localized.

To consider the source localization, we note that the intensity, I , of the hot emission can be written approximately in terms of the solar flux F_s ($\sim 250 \text{ kR}$)¹⁵, the resonance scattering cross-section σ_{rs} ($\sim 2.5 \times 10^{-13} \text{ cm}^2$) and the hot hydrogen density N_H

$$I = F_s \sigma_{rs} \int_{z_e}^{z_p} dz N_H(z) \quad (2)$$

where z_e is the altitude of the equatorward intercept and z_p is the altitude of the poleward intercept of the line of sight with the auroral flux tube. Locally, the density of hot H^0 can be parameterized in terms of a scale height h along the field line

$$N_H = N_o \exp[-(z - z_o)/h \sin \lambda] \quad (3)$$

where N_o is the density at the reference altitude z_o along the

field line at invariant latitude λ . For our case, z_p is much larger than z_e because the auroral field lines are approximately radial, yielding

$$I \approx F_s \sigma_{rs} h \sin \lambda N_o \exp -[(z_e - z_o)/h \sin \lambda] \quad (4)$$

Thus, the intensity is expected to be an exponential function of the equatorward intercept z_e , and the width of the emission source is expected to be determined by the projected scale height $h \sin \lambda$.

Figure 5 shows the comparison between a modelling of the intensity according to equation (4) and the actual observations along the satellite ground track. The observed width is well fit with a scale height of $\sim 1,000$ km and, indeed, the intensity varies approximately with the equatorward intercept as an inverse exponential. The observed intensity is obtained by subtracting the fitted interplanetary hot emission background¹³ from the total hot emission to obtain the magnetospheric component of the hot emission intensity. Even without assuming any longitudinal variations of the aurora, the two intensity curves shown in Fig. 4 are sufficiently similar in shape and location to support our conjecture and to explain the most significant feature that the magnetospheric hot emission source appears to be localized. Furthermore, using the peak intensity of ~ 200 R and the fitted

width of $\sim 1,000$ km, we can determine from equation (4) that the hot hydrogen source density is $\sim 30 \text{ cm}^{-3}$. This number density is quite reasonable, as the hot H^0 includes all populations above ~ 0.1 eV.

We thank J. L. Bertaux for providing information on the Spacelab 1 observations.

Received 20 March; accepted 29 April 1986.

1. Chakrabarti, S., Paresce, F., Bowyer, S., Chiu, Y. T. & Aikin, A. *Geophys. Res. Lett.* **9**, 151-154 (1982).
2. Chiu, Y. T., Luhmann, J. G., Ching, B. K. & Boucher, D. J. *Jr J. geophys. Res.* **84**, 909-916 (1979).
3. Shelley, E. G. *Space Sci. Rev.* **23**, 465-497 (1979).
4. Chiu, Y. T., Cornwall, J. M., Fennell, J. F., Gorney, D. J. & Mizera, P. F. *Space Sci. Rev.* **35**, 221-257 (1983).
5. Moore, T. E. *Rev. Geophys. Space Phys.* **22**, 264 (1984).
6. Johnson, R. G. *Energetic Ion Composition in the Earth's Magnetosphere* (Reidel, Dordrecht, 1983).
7. Kumar, S., Chakrabarti, S., Paresce, F. & Bowyer, S. *J. geophys. Res.* **88**, 9271-9279 (1983).
8. Bowyer, S., Kimble, R., Paresce, F., Lampton, M. & Penegor, G. *Appl. Opt.* **20**, 477 (1981).
9. Chiu, Y. T. & Schulz, M. *J. geophys. Res.* **83**, 629-642 (1978).
10. Hoffman, J. H., Dodson, W. H., Lippincott, C. R. & Hammack, H. D. *J. geophys. Res.* **79**, 4246-4251 (1974).
11. Paresce, F., Chakrabarti, S., Boyer, S. & Kimble, R. *J. geophys. Res.* **88**, 4905-4910 (1983).
12. Chakrabarti, S., Paresce, F., Bowyer, S. & Kimble, R. *J. geophys. Res.* **88**, 4898 (1983).
13. Chakrabarti, S., Kimble, R. & Bowyer, S. *J. geophys. Res.* **89**, 5660-5664 (1984).
14. Bertaux, J. L., Goutail, F. & Kockarts, G. *Science* **225**, 174-176 (1984).
15. Heroux, L. & Hinteregger, H. E. *J. geophys. Res.* **83**, 5305-5308 (1978).

Periodicity of the Earth's magnetic reversals

Richard B. Stothers

National Aeronautics and Space Administration, Goddard Space Flight Center, Institute for Space Studies, 2880 Broadway, New York New York 10025, USA

Reversals of the Earth's magnetic field may occur with a certain regularity. Using observed percentages of normal and reversed polarity during different time intervals, Negi and Tiwari¹ detected several significant periodicities, notably one of 32-34 Myr. Raup² used the dates of individual reversals to propose a 30-Myr period in the frequency of reversals, although Mazaud *et al.*³ obtained a periodicity of only 15 Myr. Like Negi and Tiwari's¹ other detected short periods, the 15-Myr period may be harmonically related to a basic 30-Myr period⁴. It has also been suggested that these are accidental periodicities arising in a short record^{5,6}, or are harmonics of the record length itself⁷. In fact, all of the cited studies have used different data, different record lengths and different methods of time-series analysis. Raup⁸ has consequently retracted his original claim. Here I present a much fuller analysis of the reversal record and show that a statistically significant period of ~ 30 Myr does formally exist, in spite of the cited differences.

For consistency with the most recent studies, most of the calculations reported here have used the dates given by Harland *et al.*⁹ for 296 magnetic reversals in the past 165 Myr. Raup² and Lutz⁷ displayed histograms of these dates in bins with widths of 5 and 8.27 Myr, respectively. Because 5 Myr is a harmonic of the suggested 30-Myr period, and 8.27 Myr is too long to reveal visually a 30-Myr period, an unbiased bin width of 4 Myr is used here to construct a new histogram, shown in Fig. 1. Inspection suggests that peaking of the magnetic reversal rate near both ends of the record will produce a very strong artificial periodicity of ~ 140 Myr, while the complete absence of magnetic reversals near the middle of the record (Cretaceous quiet interval) will create a strong first harmonic of ~ 70 Myr. Superimposed on the bowl-shaped histogram, however, there appears to be a possibly real oscillation of minor peaks with a period of ~ 30 Myr. Other published displays of the reversal frequency suggest a shorter wavelength of 15 or 20 Myr (refs 3, 10), but only the 30-Myr oscillation appears when the data are binned

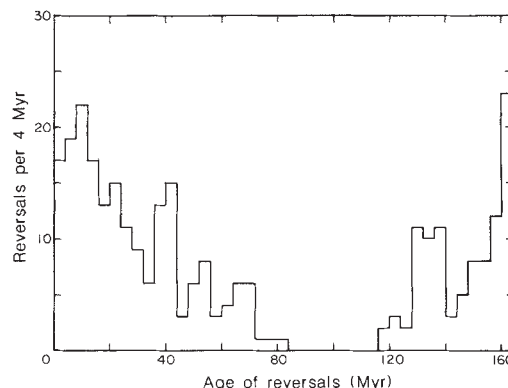


Fig. 1 Magnetic reversal frequencies shown in bins of width 4 Myr (data from ref. 9). Note the minor peaks spaced at ~ 30 -Myr intervals.

in very short intervals, such as 1 Myr.

To proceed further, the dates of individual magnetic reversals may be analysed by using a technique in which the observed dates t_i ($i = 1, 2, \dots, N$) are fitted to a linear periodic function of the type $t = t_0 + nP$, where P is a trial period, t_0 is a trial value for the most recent epoch, and n is an integer. In the original formulation of the technique¹¹, I minimized $N\sigma^2$, the sum of the squares of the residuals for each trial epoch, to obtain a best fit for a selected trial period, and computed a 'residuals index' $(\sigma - \sigma_c)/P$, where $\sigma_c/P = [(N^2 - 1)/12N^2]^{1/2}$. Raup and Sepkoski¹², in independent work, searched for the arithmetic mean of the residuals that was closest to zero, and Lutz⁷ made a circular transformation to obtain phases of the observed data, the dispersion of which was then minimized by performing two trigonometric summations, as in Fourier analysis. Lutz's variation of the method allows a correction to be made for the unequal weighting of the dates which biases most, if not all, time-series analyses. However, as he showed that this effect is very small for the present data, I neglect it here and use the technique as originally formulated. Earlier versions, for both temporal¹³ and spatial¹⁴ problems, looked for either a least-squares or maximum-likelihood solution for two unknowns and so did not compute the spectral information, or else assumed a known value for the most recent epoch.

The computed spectrum of the residuals index for the adopted dates of magnetic reversals (Fig. 2) displays many periodicities.

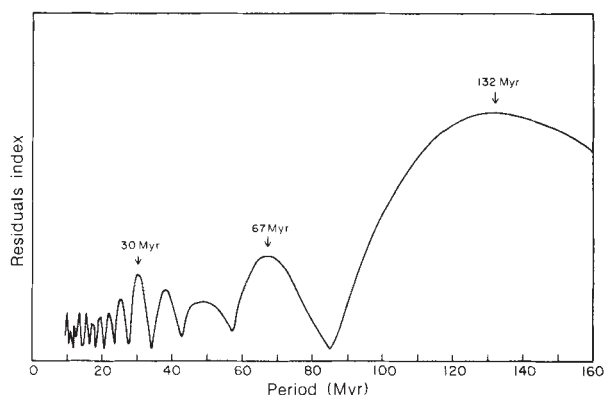


Fig. 2 Spectrum of residuals index for the complete time series of magnetic reversals, $t = 0-165$ Myr. The residuals index measures the goodness-of-fit of the observed time series to an assumed, perfectly periodic time series.

To determine their robustness, the record can be progressively truncated: successive cuts of 5 Myr are applied from the present time to as far back as 65 Myr ago (beginning of the Cenozoic Era). Table 1 lists the periods of all spectral peaks appearing at trial periods $P \geq 10$ Myr. A number of important features of this table should be noted. First, the longest period is always of the order of the length of the truncated record. (This period always has the highest spectral peak.) Second, the next-longest period is always nearly equal to one-half of the longest period. (This first-harmonic period has the next-highest spectral peak if the amount of truncation is < 25 Myr.) Third, the dominant period occurring elsewhere in the computed spectra is 28-33 Myr. Its precise value oscillates down Table 1 in a discontinuous, saw-tooth fashion with a wavelength of ~ 30 Myr. The same wavelength of oscillation (with a slight phase shift) characterizes the neighbouring periods of 22-25 and 20-22 Myr, although the shorter periods of 18-19, 17-18 and 15-16 Myr vary down the table with a wavelength only half as large. Note that the percentage by which the period varies is smaller for the shorter periods, so that the variations of the four shortest periods, 13, 12, 11 and 10 Myr, are imperceptible. Thus, the more cycles a record contains, the less these cycles will be distorted by the record length.

The average values of the 12 periods that lie in the range 10-58 Myr are closely represented by the individual periods shown in the first row of Table 1, and are very close to what would be expected for an ideal time series consisting of six dates successively separated by 30.1 Myr. An explicit comparison of observed and expected periods is presented in Table 2, including a simple harmonic interpretation¹¹. The correlation coefficient between the 11 pairs of observed and expected periods (excluding the generating period of 30.1 Myr) is $r = 0.997$. The significance of this value has been tested by correlating the sequence of 11 observed periods with 5,000 simulated sequences, each consisting of 11 numbers drawn at random from the range 10-58 Myr and then ordered. Fewer than 0.1% of the simulated sequences showed $r \geq 0.997$. Since there is no *a priori* reason to expect precisely 11 periods to show up in the range 10-58 Myr, this represents a very conservative test of significance.

An independent test has been made by generating 5,000 artificial time series, each containing 296 dates selected randomly from the time interval 0-165 Myr and then ordered. The spectrum of the residuals index has been computed for each time series and compared with Fig. 2. Fewer than 0.1% of the synthetic spectra showed a residuals index at $P = 30$ Myr that was higher than the observed peak. However, this test assumes 30 Myr as an *a priori* period and does not take into account the non-stationarities in the observed time series, as illustrated in Fig. 1.

Raup² performed a different significance test by randomizing the order of the observed time intervals. He found a low percentage (0.6%) of simulated cases in which the height of the observed spectral peak at $P = 30$ Myr was surpassed. Lutz⁷ detected the statistical significance of the observed peaks at shorter periods, but did not recognize their harmonic connection to the 30-Myr period. Instead, he associated these periods with the record length, an interpretation that was based on a limited truncation back to only 21 Myr BP. Additional evidence against Lutz's interpretation follows from an attempt to match the provisionally expected series of harmonics ($1/2, 1/3, 1/4, \dots$) of the longest period in each row of Table 1 to the rest of the observed periods along the same row. Discrepancies of up to 15% in the values of the observed periods occur, and not all of the observed periods can be successfully identified. The number of extraneous periods per row is three in one of the rows, two in six rows, and one in another six rows; only one row has no extraneous period at all.

Truncation of the record in the middle of the Cretaceous quiet interval provides a further test. Separate spectral analyses of the more recent time segment ($t = 0-83$ Myr, $N = 196$ dates) and the older time segment ($t = 118-165$ Myr, $N = 100$ dates) yield the following results for trial periods in the range 10-40 Myr. The more recent time segment contains five spectral peaks at periods of 32.7, 20.0, 15.9, 13.3 and 9.8 Myr; the older time segment contains three peaks at 29.9, 14.2 and 10.4 Myr. For comparison, two ideal time series consisting of three dates and two dates, respectively, with generating periods of 32.7 Myr and 29.9 Myr,

Table 1 Periods of the spectral peaks in time-series analyses of a progressively truncated record of magnetic reversals

t	Periods (Myr)													
0-165	132	67	48	38	30*	25	22	19	17	16	13	12	11	10
5-165	128	65	48	37	29*	24	21	19	17	15	13	12	11	10
10-165	123	63	47	36	29*	24	20	18	17	15	13	12	11	10
15-165	117	60	46	33	28*	23	20	—	18	16	13	12	11	10
20-165	113	59	44	33*	28	22	20	—	17	15	13	12	11	10
25-165	109	57	42	32*	—	25	22	19	17	15	13	12	11	10
30-165	105	55	41	30*	—	24	21	—	18	16	13	12	11	10
35-165	102	52	40	29*	—	23	20	—	17	15	13	12	11	10
40-165	99	51	38	29*	—	23	20	19	17	15	13	12	11	10
45-165	96	49	34	28*	—	22	—	—	18	16	13	12	11	10
50-165	94	48	33*	28	—	22	—	—	18	15	13	12	11	10
55-165	93	47	32*	—	—	24	—	19	17	15	13	12	11	10
60-165	93	45	31*	—	—	23	—	19	—	16	13	12	11	10
65-165	96	43	30*	—	—	23	—	18	—	15	13	12	11	10

* Highest spectral peak for periods < 58 Myr.

show high peaks only at 32.7, 21.8, 16.3, 13.1 and 10.9 Myr and only at 29.9, 14.9 and 10.0 Myr. The corresponding harmonic identifications are 1, 2/3, 1/2, 2/5 and 1/3, and 1, 1/2 and 1/3.

Several conclusions can be drawn from this comparison. One is that the 15-Myr period detected by Mazaud *et al.*³ in the more recent time segment is probably just a harmonic period, rather than the basic period. Although progressive truncation of this segment of the record back to 35 Myr BP leads, in some cases, to the highest spectral peak being at $P = 13-15$ Myr, comparison with the computed spectra of the corresponding ideal time series shows unambiguously that the generating period remains $P \approx 30$ Myr (plus or minus the amount of period oscillation). Second, the presence of the 30-Myr period in the separate spectral analyses for $t = 0-83$ Myr and $t = 118-165$ Myr demonstrates that this periodicity is not being produced by the 35-Myr quiet interval. (Raup² came to the same conclusion.) Third, the abrupt shifts in the number of observed periods, going down Table 1, arise from the change in the number of cycles covered as the record is progressively truncated. For an ideal time series, the total number of integer multiples of harmonics that appear in

Table 2 Periods (Myr) of the peaks in spectral analyses of the magnetic-reversal and ideal time series

Observed series	Ideal series	Harmonic series	Harmonic series identification
48.2	48.0	45.2, 50.2	3/2, 5/3
38.0	40.5	40.1, 37.6	4/3, 5/4
30.1	30.1	30.1	1
25.2	24.0	24.1, 25.1	4/5, 5/6
—	22.0*	22.6, 21.5	3/4, 5/7
21.9	20.7	20.1, 21.5	2/3, 5/7
19.4	19.5	20.1, 18.8	2/3, 5/8
—	18.5*	18.1, 18.8	3/5, 5/8
17.2	17.3	17.2	4/7
15.5	15.1	15.1	1/2
13.4	13.1	12.9	3/7
12.0	12.0	12.0	2/5
10.9	11.0	11.3	3/8
10.0	10.0	10.0	1/3

* Very low spectral peak.

the computed spectrum is by definition controlled by the value of the largest-occurring integer, which turns out to be equal to the number of cycles covered by the time series.

Three additional checks have been made of the above results. First, Fourier analysis has been applied to the histogram shown in Fig. 1. Spectral peaks were determined to occur with relative heights and absolute periods that agreed, within the resolution of the method, with the peaks found by linear analysis. To examine the effect of the obvious non-stationarities in Fig. 1, de-trending was performed by subtracting the strongest contributing Fourier components. Accordingly, the two longest cycles, corresponding to the 132-Myr and 67-Myr periods of Fig. 2, were successively removed. Fourier analysis of the residual data then showed the highest spectral peak occurring at 30.9 Myr, which was shifted only slightly from the original value of 30.4 Myr. A similar Fourier analysis of Raup's² and Lutz's⁷ histograms, as well as of a histogram with a 1-Myr bin width, yielded essentially identical results. This check shows that the method of analysis does not bias the results.

A second test concerned the possible influence of undetected magnetic reversals. If the unknown reversals are distributed randomly or proportionally among the known ones, their omission should not significantly affect the present results. To simulate their omission, the known magnetic reversals were systematically pruned from the record. Only the recent time segment $t = 0-83$ Myr was considered, and all fixed-polarity chrons of less than a certain length δt were excluded from the data set. As the missing chrons from the original record are probably shorter than 0.04 Myr (ref. 10), a very conservative test uses $\delta t \gg 0.04$ Myr. When $\delta t = 0.5$ Myr is used, the basic period of 30.1 Myr still emerges ($N = 49$ dates). Even with $\delta t = 1$ Myr, the period remains 28.9 Myr ($N = 19$ dates).

A third check of the present results involved using two different sets of magnetic-reversal dates: those of Kent and Gradstein¹⁵ ($t = 0-169$ Myr, $N = 283$ dates) and of Lowrie and

Alvarez¹⁶ ($t = 0-84$ Myr, $N = 174$ dates). Analysis of these dates by the linear technique produced the results shown in Table 3. In view of the small number of cycles covered, the differences seem rather minor, the total spread among the periods being ± 3 Myr. Using maximum-entropy analysis, Pal and Creer¹⁷ independently found the 32.1-Myr period in the data of ref. 16.

The above results suggest that a real periodicity exists in the record of magnetic reversals, supporting the claims of Negi and Tiwari¹ and of Raup². With fine tuning of the truncated-record results, the mean value of the basic period falls at 30.5 Myr (ref. 9 data) or 29.5 Myr (ref. 15 data). Shuter and Klatt's¹⁸ 136-Myr period is now seen to arise only from the record length. Like the basic period, the phase oscillates in a discontinuous, saw-tooth fashion with progressive truncation, and returns to its starting value every ~ 30 Myr. A best estimate of its mean value is $t_0 = 10 \pm 8$ Myr BP (ref. 9 data) or $t_0 = 7 \pm 8$ Myr BP (ref. 15 data). Note that t_0 represents the most recent epoch of the mean cycle and therefore is not necessarily equal to the most recent time of high magnetic-reversal frequency, unless the times are strictly periodic.

These results do not prove that there is a physical periodicity in the Earth's magnetic record, nor do they prove that any periodicity, real or accidental, is strictly regular. However, the already demonstrated^{4,19} coincidences in time between a high rate of magnetic reversals, heightened global tectonic activity, and impact cratering do at least suggest that periodic or episodic large-body impacts may trigger many of these disturbances.

I thank D. V. Kent, T. M. Lutz, P. L. McFadden, M. R. Rampino and D. M. Raup for helpful communications.

Received 9 December 1985; accepted 7 May 1986.

1. Negi, J. G. & Tiwari, R. K. *Geophys. Res. Lett.* **10**, 713-716 (1983).
2. Raup, D. M. *Nature* **314**, 341-343 (1985).
3. Mazaud, A., Laj, C., de Sèze, L. & Verosub, K. L. *Nature* **304**, 328-330 (1983).
4. Rampino, M. R. & Stothers, R. B. *Science* **226**, 1427-1431 (1984).
5. McFadden, P. L. *Nature* **311**, 396 (1984).
6. Mazaud, A., Laj, C., de Sèze, L. & Verosub, K. L. *Nature* **311**, 396 (1984).
7. Lutz, T. M. *Nature* **317**, 404-407 (1985).
8. Raup, D. M. *Nature* **317**, 384-385 (1985).
9. Harland, W. B. *et al.* *A Geological Time Scale* (Cambridge University Press, 1982).
10. Lowrie, W. & Kent, D. V. *Earth planet. Sci. Lett.* **62**, 305-313 (1983).
11. Stothers, R. *Astr. Astrophys.* **77**, 121-127 (1979).
12. Raup, D. M. & Sepkoski, J. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 801-805 (1984).
13. Fraser, D. A. S. *Statistics: An Introduction*, 245 (Wiley, New York, 1958).
14. Broadbent, S. R. *Biometrika* **42**, 45-57 (1955); **43**, 32-44 (1956).
15. Kent, D. V. & Gradstein, F. M. in *The Western Atlantic Region, Volume M of the Geology of North America* (eds Tucholke, B. E. & Vogt, P. R.) (Geological Society of America, Boulder, in the press).
16. Lowrie, W. & Alvarez, W. *Geology* **9**, 392-397 (1981).
17. Pal, P. C. & Creer, K. M. *Nature* **320**, 148-150 (1986).
18. Shuter, W. L. H. & Klatt, C. *Astrophys. J.* **301**, 471-477 (1986).
19. Rampino, M. R. & Stothers, R. B. *Nature* **308**, 709-712 (1984).

Thermohaline intrusions created isopycnally at oceanic fronts are inclined to isopycnals

J. D. Woods*, R. Onken & J. Fischer

Institut für Meereskunde an der Universität Kiel, FRG

Table 3 Summary of results for the basic period (P) and phase (t_0) of magnetic reversals

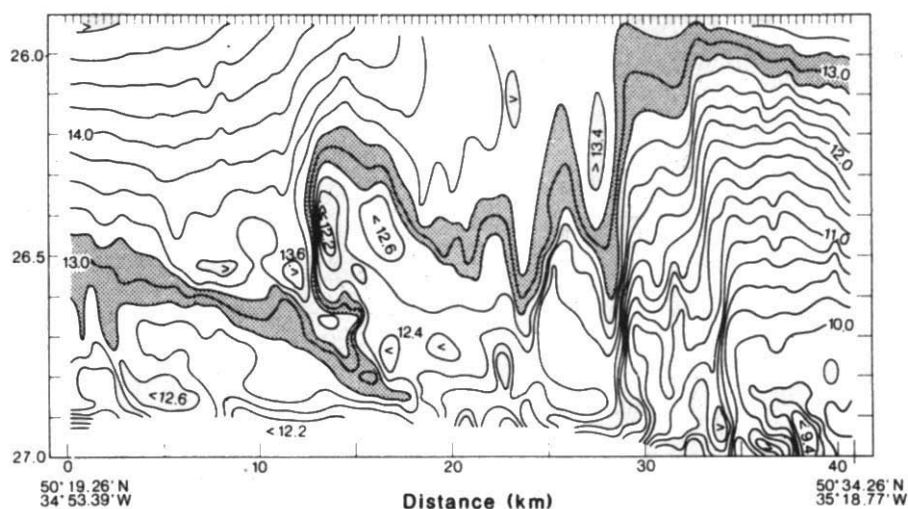
Time range (Myr)	Ref. 9 dates		Ref. 15 dates		Ref. 16 dates	
	P (Myr)	t_0 (Myr BP)	P	t_0	P	t_0
(0-65)-170*	30.5	10	29.5	7	—	—
0-170	30.1	10	29.9	7	—	—
(0-35)-85*	28.0	14	29.0	10	28.0	14
0-85	32.7	9	28.8	7	32.1	9
115-170	29.9	10	26.9	-6	—	—

* Average of the truncation results.

Thermohaline intrusions have often been observed extending several kilometres from ocean fronts¹. They have a central role in models of mixing^{2,3}; however, their origin has been the subject of controversy. Laboratory studies led to the idea that intrusions might be formed by double diffusion⁴⁻⁶, in which case they would slope across isopycnals (surface of constant density). That signature has been identified in ocean fine-structure sections⁷. Observations of the three-dimensional structure of meandering fronts led to a different conjecture: that intrusions might be formed isopycnally by the ageostrophic circulation within unstable meanders on the frontal jet^{1,8}. In that conceptual model, double diffusion

* Present address: NERC, Polaris House, North Star Ave, Swindon, Wiltshire SN2 1EU, UK.

Fig. 1 Density variation (σ_t) of isotherms in a vertical section across the North Atlantic polar front, surveyed with the *Kiel Sea Rover* apparatus in summer 1981. The horizontal resolution (400 m) is indicated by the tick marks on the upper axis; the vertical depth range lies between 30 and 80 m. The temperature range between the 12.8 °C and 13.2 °C isotherms is shaded.



does not initiate intrusions, but may extend and dissipate those created by frontal circulation⁹. Here we present first the observed structure of such an intrusion and then the results of a dynamical model simulating the isopycnic generation mechanism at an unstable meandering front. We show that thermohaline intrusions formed isopycnically also slope across density surfaces, so there is no need to invoke double diffusion to explain such structures.

We have observed a thermohaline intrusion at a meander structure of the North Atlantic polar front. The section presented here (Fig. 1) was obtained by the *Kiel Sea Rover* towed CTD (conductivity/temperature/depth) system¹⁰. It represents a high-resolution snapshot of a thermohaline front in the seasonal pycnocline. The orientation of the section was almost perpendicular to the front which separates the warm and saline water in the anticyclonic meander-ridge, from the cold and fresh water in the cyclonic trough of the meander. Isopycnic analysis eliminated internal wave noise¹¹. The resulting product revealed strong thermohaline fine structure created by the three-dimensional frontal dynamics.

The 13 °C isotherm is folded in the vicinity of the major thermocline signal. Relatively warm water is intruded into a region of cooler and fresher water, with the axis of the intrusion being clearly inclined to isopycnic surfaces. The horizontal extent of the fold was ~6 km and the vertical scale of the intrusion, estimated from the mean static stability profile, was ~10 m.

We modelled the development of a meandering front in two stages: (1) frontogenesis, which transforms the initial (large-scale) baroclinicity and thermoclinicity into a jet with an associated $T-S$ (temperature-salinity) front, and (2) instability of the jet. It is important that the initial condition for stage (2) is the output of stage (1), so that we can establish the sensitivity of intrusion structure to large-scale initial conditions. Our simulation of thermohaline intrusions was based on a primitive equation model with quasi-isopycnic coordinates¹². The first stage, describing frontogenesis in two dimensions, was an extension of that described in ref. 9, with initial thermoclinicity described by horizontal isotherms. It gave the frontal structure shown in Fig. 2a. That was used as the initial conditions for the second stage¹³, a channel model based on the Bleck-Boudra quasi-isopycnic coordinate scheme¹², rigid side walls with free slip, and cyclic downstream boundary conditions. The grid has 10 isopycnic layers, each with a rectangular array of 32×64 grid points, spaced 2.5 km apart in the cross-stream direction and in the downstream direction. Meanders grow in response to an initial small perturbation in the stream function. In the present examples, the disturbance had downstream amplitude as shown by the streamlines at a depth of 25 m after 40 days of integration (Fig. 3a). The corresponding pattern of isotherms on an isopycnic surface (Fig. 3b) appears much more complicated, but is,

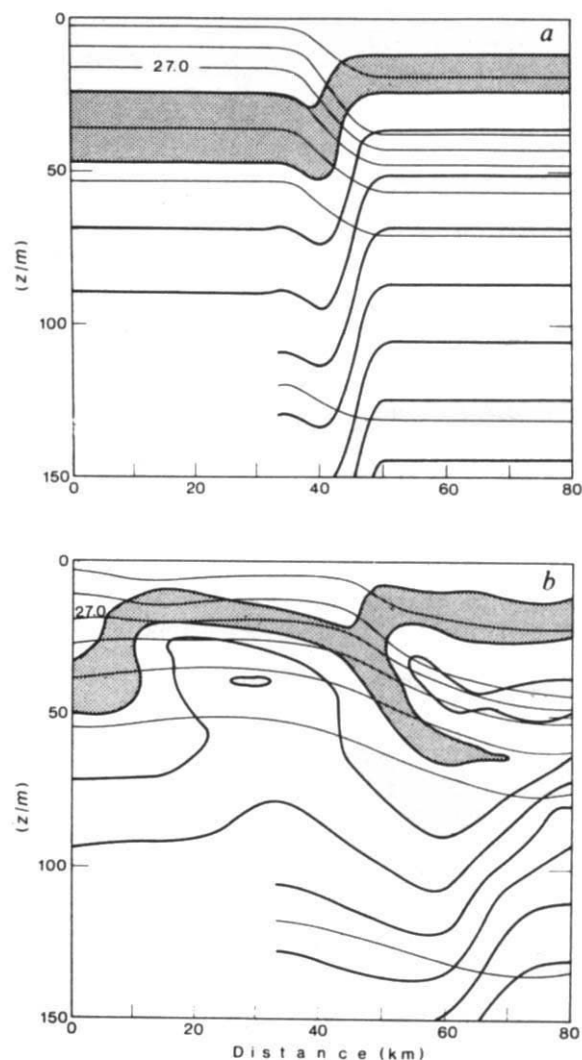


Fig. 2 Depth of isotherms (thick contours) and isopycnals (thin contours) in a vertical section of the upper 150 m in the model domain at positions indicated by the broken lines in Fig. 3. The temperature range between the 18 °C and 19 °C isotherms is shaded. a, Initial conditions; b, day 40.

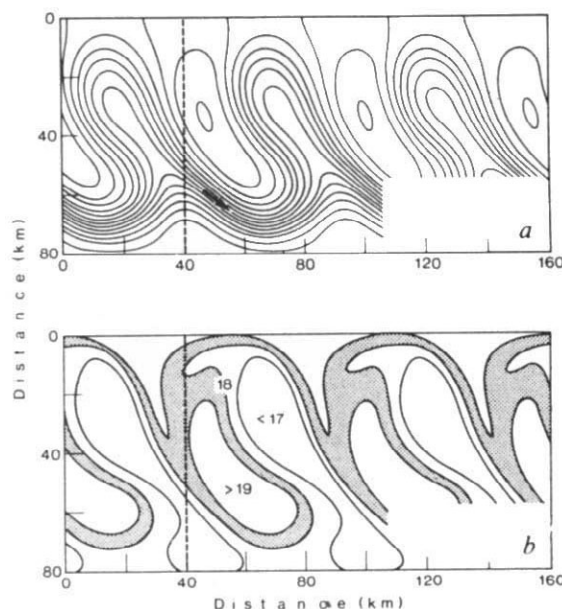


Fig. 3 a, Stream function at 25 m depth after 40 days of integration of the frontal meander model. The arrow indicates the flow direction. b, Temperature pattern on the isopycnal surface $\sigma_t = 27.0 \text{ kg m}^{-3}$ at day 40. The mean depth of this surface is $\sim 25 \text{ m}$.

of course, consistent with the evolving streamlines. A vertical section through the meander (Fig. 2b) shows an intrusion of relatively warm, salty water into the cooler, fresher side of the front. Its axis is inclined to isopycnals. The 18–19 °C temperature range has a 20-km fold.

To understand the isopycnal mechanism of thermohaline intrusions at fronts, it is necessary to discriminate between kinematic and dynamic processes. The latter change the pressure gradient and accelerate the flow; the former do not. Kinematic frontogenesis is the result of a horizontal confluence acting on a passive scalar gradient, which becomes sharpened without changing the flow. The passive gradient of interest here is thermoclinicity, defined as the gradient of temperature on a surface of constant density. Dynamical frontogenesis is the result of a horizontal confluence acting on a dynamically active scalar gradient, which changes the flow as it becomes sharper. The active scalar gradient of interest here is baroclinicity, defined as the gradient of pressure along a surface of constant density. Both processes occur together where there is horizontal confluence of a body of water that is both thermoclinic and baroclinic; that is, one in which the isothermal surfaces are inclined to non-horizontal isopycnal surfaces—the necessary condition for the isopycnal mechanism of thermohaline intrusions. It depends on the kinematic folding of an isothermal surface by the total velocity field, comprising the externally imposed confluence plus the internal circulation arising from the dynamic response to that confluence.

Regional variation of air-sea interaction creates large-scale variations of temperature and salinity in the surface mixed layer. After subduction¹⁴ into the interior these large-scale variations include both baroclinicity and thermoclinicity. Frontogenesis occurs where there is a confluence in the (mostly laminar-flow¹⁵ and therefore isopycnal) circulation. The thermoclinicity is sharpened kinematically into a T - S front. The corresponding sharpening of the baroclinicity produces a jet, which is roughly in geostrophic balance with the density gradient. To simulate the cross-front shear that produces intrusions, we have to compute the ageostrophic component of flow associated with the acceleration of the jet. That is done by using the law of isopycnal potential vorticity (IPV) conservation, which states that $IPV = ((\zeta + f)/h)(d/D) = \text{constant}$, for a water column bounded by

a pair of isopycnals, where f is the planetary vorticity, and h is the spacing between a pair of isopycnals with mean density D and density difference d . According to this law, h varies with the absolute vorticity, $(\zeta + f)$. Relative vorticity of large-scale motion is negligible compared with f , so the initial IPV before frontogenesis is given by $IPV = (f/h(0))/(d/D)$. As frontogenesis proceeds, changes in h are computed from the local Rossby number (Ro), $h(t)/h(0) = \zeta(t)/f = Ro(t)$. In practice, the jet scale is usually so small that changes in f can be neglected in comparison with changes in ζ . But as the jet accelerates, the cross-jet shear begins to develop significant relative vorticity (ζ). The spacing between isopycnals adjusts accordingly, becoming greater on the cyclonic side, and smaller on the anticyclonic side. Continuity of mass is achieved by a cross-jet, ageostrophic flow¹⁶. The structure of this cross-jet circulation depends on the initial profile of the isopycnal gradient of IPV (IPVG)¹⁷. The spring subduction process tends to create a maximum of IPVG near the top of the seasonal thermocline¹⁸. In that case the cross-front ageostrophic circulation comprises a simple cell, with upwelling on the anticyclonic side and downwelling on the cyclonic side of the jet. The circulation tends to flatten isopycnals, limiting the final kinetic energy of the jet, preventing further narrowing and progressively decreasing its depth. The kinematic sharpening of frontal thermoclinicity, effected mainly by the external confluence, is modulated by divergence of the cross-jet ageostrophic circulation. It does not become as sharp as it would have done due to the confluence alone, but, nevertheless, ends up much narrower than the jet. The surface of thermoclinicity maximum, which may have been initially vertical, becomes tipped over, so that it slopes down from the cyclonic side to the anticyclonic, creating the well-documented slope of the temperature field at a front¹⁹. These characteristics (narrowness and slope of the thermoclinicity maximum relative to the baroclinicity maximum) are important for the structure of thermohaline intrusions which develop as the jet begins to meander unstably.

Frontogenesis, as described above, leads to a two-dimensional structure which can be used as initial conditions for modelling jet instability. The jet is normally unstable, with fastest-growing wavelength and growth rate dependent on the initial IPVG profile. As the meanders grow to large amplitude, they develop curvature vorticity which modulates the spacing between isopycnals, according to the law of IPV conservation described above. Ageostrophic flow from thinning regions to thickening regions advects isotherms isopycnally across the meandering jet core, with a complex vertical structure governed by the initial conditions (and ultimately by the profile of large-scale IPVG before frontogenesis). At any moment, the pattern formed by isotherms on each isopycnal surface depends on the initial thermoclinicity (before frontogenesis) and the time integral of the total flow, including both the geostrophic jet component and the ageostrophic vortex-stretching component.

Our model proves that the isopycnal motion in unstable fronts can produce thermohaline intrusions that slope across isopycnals, a property previously believed to be symptomatic of the double diffusion mechanism. It also emphasizes the three-dimensional character of the velocity field responsible for isopycnal intrusions: one must resist the temptation to seek two-dimensional explanations for intrusions observed in hydrographic sections. Finally, we emphasize that the structure shown in Figs 2 and 3 depended critically on our choice of initial conditions. Cross-isopycnal folding of isotherms is common, but the details vary. The strength of our model technique is that it permits us accurately to simulate the complicated development from explicit large-scale initial conditions to a meandering jet/front with high Rossby number.

Received 7 March; accepted 29 April 1986.

1. Woods, J. D., Wiley, R. L. & Briscoe, M. G. *Deep-Sea Res.* **24**, 253–275 (1977).
2. Joyce, T. J. *phys. Oceanogr.* **7**, 626–629 (1977).
3. Garrett, C. J. *phys. Oceanogr.* **12**, 952–957 (1982).

4. Turner, J. S. *J. geophys. Res.* **83**, 2887-2901 (1978).
5. Turner, J. S. in *Evolution of Physical Oceanography* (ed. Warren, B. A. & Wunsch, C.) 236-262 (MIT, Cambridge, 1981).
6. Ruddick, B. R. & Turner, J. S. *Deep-Sea Res.* **903**-913 (1979).
7. Gargett, A. E. *J. geophys. Res.* **83**, 5123-5134 (1978).
8. Gregg, M. G. & McKenzie, J. H. *Nature* **280**, 310-311 (1979).
9. Woods, J. D. *Ocean Model.* **32**, 1-4 (1978).
10. Bauer, J., Fischer, J., Leach, H. & Woods, J. D. *Ber. Inst. Meeresk. Univ. Kiel, FRG*. No. 143 (1985).
11. Leach, H., Minnett, P. J. & Woods, J. D. *Deep-Sea Res.* **32**, 575-597 (1985).
12. Bleck, R. & Boudra, D. B. *J. phys. Oceanogr.* **11**, 755-770 (1981).
13. Onken, R. thesis Univ. Kiel (1986).
14. Woods, J. D. & Barkmann, W. *Q. Jl R. met. Soc.* **112**, 1-27 (1986).
15. Woods, J. D. *J. Fluid Mech.* **32**, 791-800 (1986).
16. Sawyer, J. S. *Proc. R. Soc. A234*, 364-362 (1956).
17. Bleck, R., Onken, R. & Woods, J. D. *Q. Jl R. met. Soc.* (1986).
18. Stammer, D. & Woods, J. D. *Ber. Inst. Meeresk. Univ. Kiel, FRG* (in the press).
19. MacVean, M. K. & Woods, J. D. *Q. Jl R. met. Soc.* **106**, 293-311 (1980).

Effects of surface heat flux during the 1972 and 1982 El Niño episodes

R. K. Reed

Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, Washington 98115, USA

It has been suggested^{1,2} that anomalous surface heat flux may be partly responsible for the initial warming that occurs in the eastern Pacific during El Niño events. Sea surface temperature, net surface heat flux, and winds in the equatorial Pacific are examined here for the 1972 and 1982 El Niño episodes. It is found that surface heat flux and sea surface temperature anomalies tend to have little or negative correlation; thus, surface flux is not of major importance in the formation of thermal anomalies in the central and eastern Pacific.

In northern summer of 1982 a large El Niño episode began. This event produced dramatic effects in the equatorial Pacific Ocean (disappearance of the equatorial undercurrent³, for example) and induced global changes in atmospheric conditions⁴. Initially, there was considerable discussion of the differences between this event and a 'canonical' episode. Further examination of some of these differences, however, revealed that they largely resulted from similar processes acting in the presence of different initial conditions⁵. It would be of interest to explore in some detail the morphology of the 1982 episode in comparison with a more typical event such as that of 1972. Special emphasis is placed on examining surface heat flux, in an effort to resolve its role in creating thermal anomalies.

The data used for this study were archived ship-of-opportunity weather reports. They were obtained from NOAA's National Climatic Data Center for the periods 1970-79 and January 1982-October 1983. The 1970-79 period was chosen primarily because this was the first data subset for which careful editing procedures had been used to eliminate duplicate reports and other erroneous data. The climatology during this decade was also found to agree closely with those based on longer periods⁶. Surface heat fluxes (insolation, net long-wave radiation, latent heat, sensible heat, and their sum—the net or total heat flux) were calculated from the individual weather reports using empirical formulas⁶, and the fluxes were averaged over 2° latitude by 10° longitude areas to obtain monthly values. Weather elements were also averaged to obtain monthly means.

Estimates of the standard errors of the mean values for each month were made from the calculated intra-month variances for the 2°×10° areas used below. Because the data are very heterogeneous, the estimates are biased high by natural spatial and temporal variability, but they do provide conservative guides. Keeping these caveats in mind, standard errors of the mean sea surface temperature and net heat flux, for an individual month, were calculated as 0.7 °C and 45 W m⁻².

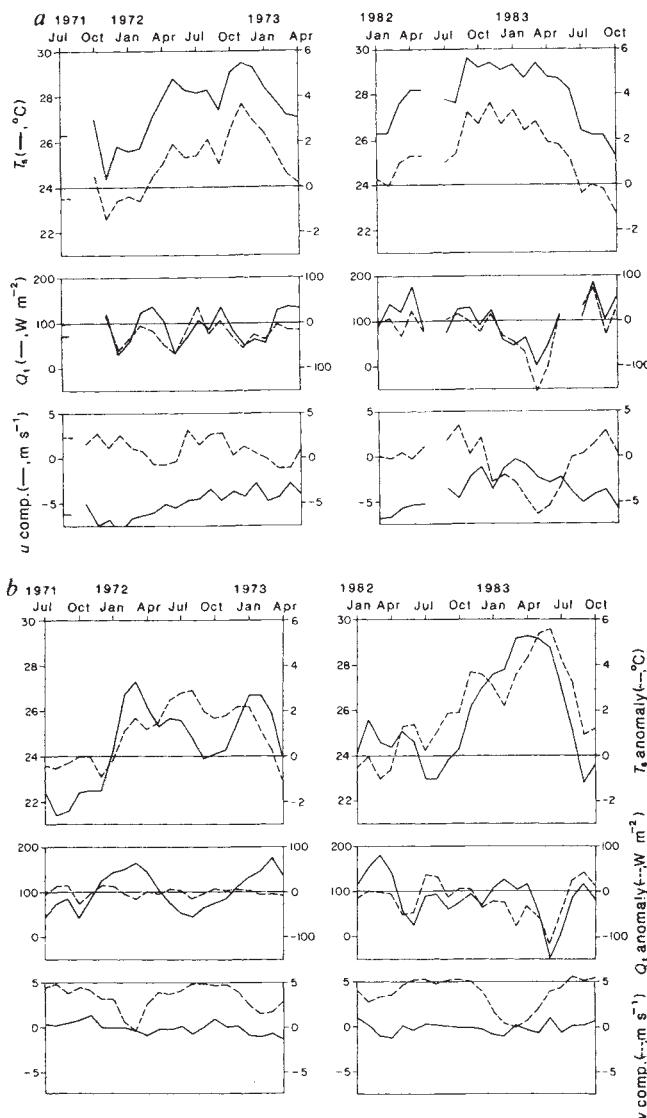


Fig. 1 *a*, Plots of mean monthly sea surface temperature (T_s , °C), sea surface temperature anomaly (individual monthly mean minus monthly mean for the period 1970-79), mean monthly net or total surface heat flux (Q_t , positive for flux into the ocean), net surface heat flux anomaly (individual monthly mean minus monthly mean for the period 1970-79), and the u (east)- and v (north)-components of wind velocity (in m s⁻¹) for the central equatorial Pacific (2° N-2° S, 140-160° W), July 1971-April 1973 and January 1982-October 1983. *b*, As *a* except for the eastern equatorial Pacific (2° N-6° S, 80-90° W).

Figure 1*a* presents information on the evolution of surface properties during both the 1972 and 1982 El Niño episodes for a region in the central Pacific (2° N-2° S, 140-160° W). When more than two of the original four 2°×10° areas had fewer than three observations, the data were omitted. Each monthly mean was typically based on >10 observations. The east (u) and north (v) components of wind are shown rather than wind speed or anomalies.

During the 1972 episode, surface temperature (T_s) anomalies increased by 2.5 °C between February and May 1972. A further increase occurred during September-November, with a rapid decrease thereafter. (All of the positive temperature anomalies between April 1972 and February 1973 are greater than the standard error estimate above.) Both warming periods coincided with consistently negative surface heat flux (Q_t) anomalies, although some of the values are smaller than the error estimate.

The u -component of wind at this site decreased in magnitude through much of the period, but the winds did not become eastward.

The 1982 event evolved differently. There was a small increase in T_s anomaly between February and April 1982 and a large increase between July and September. Relatively constant values then occurred, followed by a fairly rapid decrease after March 1983. Q_t anomalies were small before December 1982, but large negative values were present during January–April 1983. This period coincided with small u -component winds but large negative v -component winds, a very anomalous situation⁷.

An interesting difference between the events is that during 1982, formation of the anomalies in the central Pacific led similar changes at the eastern margin by one to two months (Fig. 1b); conversely, in 1972, warming in the central region seemed to generally lag that off South America. Although the magnitudes of the maximum anomalies during the two events in the central region were similar, the duration of anomalies $>2^\circ\text{C}$ was considerably longer in 1982–83 than in 1972.

Figure 1b presents data for a region just off the South American coast (2°N – 6°S , 80° – 90°W). This region has numerous ship-of-opportunity reports, and each $2^\circ \times 10^\circ$ area generally had >20 observations. In 1972, the T_s anomaly appears to have largely formed during January to March, but a subsequent increase occurred from May to August. The second pronounced maximum in T_s , but not in T_s anomaly, in January–February 1973 gives the typical double-peak signature⁸. The Q_t anomalies during this episode were all very small, and the only unusual behaviour in winds is a sharp decrease in the southerlies during February–March 1972.

The first appreciable change in T_s anomaly in 1982 was an increase of 1.9°C during April–May. (During this period of normal seasonal cooling there was only a small increase in T_s .) A second increase in T_s anomaly occurred between August and November 1982. A final warming event took place between February and May 1983, when the T_s anomaly reached $+5.4^\circ\text{C}$, and cooling started after June. Unlike the 1972 episode, Q_t anomalies were variable early in the event, but they were all negative from December 1982 to July 1983. The southerly winds were quite weak during January–May 1983, which was of longer duration, and later in the episode, than the comparable weakening during 1972.

Figure 1 shows that during both the ‘canonical’ event of 1972 and the ‘summer’ event of 1982, heat flux anomalies were small or negative during warming. Thus, surface heat fluxes do not seem to be a significant factor in forming thermal anomalies in the central and eastern ocean. Other studies^{6,9–11} provide support for this conclusion. What processes then produce the marked warming? The surface data set used here does not allow very definitive answers, but other studies (using sub-surface thermal data, sea level data, and direct current measurements) have yielded essential agreement on the important mechanisms.

The initiation of warming near the Equator in the central and eastern ocean appears to be consistently induced by heat advection resulting from eastward-moving Kelvin waves^{12,13}. The effects will appear different depending on time of year⁵. If the event is initiated early in the year, as is typical, major warming will first appear off South America because thermal gradients are quite weak along the Equator in the central ocean in southern summer. If the event is triggered in mid-year, as in 1982, warming advances eastward across the ocean in the presence of strong east–west gradients in both the central and eastern ocean. During canonical events, final increases in temperature anomaly in the central ocean may result from a weakening or southward displacement of the South Equatorial Current¹². A southward displacement of this flow¹⁴ also appears to be a factor in the rapid rise in temperature anomaly (Fig. 1b) in early 1983, although the very weak winds then probably also resulted in reduced upwelling.

I thank G. Meyers and E. Sarachik for helpful discussions. The efforts of D. G. Kachel in retrieving the data and programming computations are appreciated. This study was supported by the Equatorial Pacific Ocean Climate Studies (EPOCS) program of NOAA. Contribution 774 from NOAA/Pacific Marine Environmental Laboratory.

Received 13 February; accepted 25 April 1986.

1. Weare, B. C. *Science* **221**, 947–949 (1983).
2. Leetma, A. J. *phys. Oceanogr.* **13**, 467–473 (1983).
3. Firing, E., Lukas, R., Sadler, J. & Wyrki, K. *Science* **222**, 1121–1122 (1983).
4. Rasmusson, E. M. & Wallace, J. M. *Science* **222**, 1195–1202 (1983).
5. Harrison, D. E. & Schopf, P. S. *Mon. Weath. Rev.* **112**, 923–933 (1984).
6. Reed, R. K. *J. geophys. Res.* **88**, 9627–9638 (1983).
7. Wyrki, K. & Meyers, G. *The Tradewind Field Over the Pacific Ocean Part II*. (Univ. Hawaii Rep. HIG-75-2, 1975).
8. Busalacchi, A. & O'Brien, J. J. *J. geophys. Res.* **86**, 10901–10907 (1981).
9. Barnett, T. P. *Mon. Weath. Rev.* **112**, 2388–2400 (1984).
10. Gill, A. E. J. *phys. Oceanogr.* **13**, 586–606 (1983).
11. Gill, A. E. & Rasmusson, E. M. *Nature* **306**, 229–234 (1983).
12. Wyrki, K. *J. phys. Oceanogr.* **7**, 779–787 (1977).
13. Lukas, R., Hayes, S. P. & Wyrki, K. *J. geophys. Res.* **89**, 10425–10430 (1984).
14. Wyrki, K. *Geophys. Res. Lett.* **12**, 125–128 (1985).

Uplift of the shores of the western Mediterranean due to Messinian desiccation and flexural isostasy

Sonya E. Norman & Clement G. Chase

Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

During the Messinian Stage (5.5 Myr, Miocene/Pliocene boundary) the $4.2 \times 10^{23} \text{ m}^3$ of water that now fills the Mediterranean evaporated. Evidence for this includes palaeogorges 1 km below the present Nile and Rhone valleys and evaporite deposits, sampled by cores and deep-sea drilling^{1,2}, that are thicker than 1 km over much of the Mediterranean³. Two-dimensional flexure models, presented here, indicate that the regionally compensated crustal upwarping from removal of the seawater load would lead to a Messinian geomorphology with an uplifted Mediterranean basin and shoreline bulges along its northwestern and southeastern coasts. These shoreline bulges would cause a reversal of downhill gradient direction in areas with low original seaward slopes. Such a profile would lead to a landward reversal of drainage in rivers with low discharge.

We used a cross-section north from Annaba, Algeria to Marseilles, France to determine present-day ocean depth, which we presume to have been comparable just before the last Messinian desiccation. Vertical deflection from removal of the seawater load along this profile was calculated from an equation derived by Hetenyi⁴ for distributed block loads and modified by Jordan⁵ for an elastic lithosphere. Figure 1 shows the flexure from 250 km inland in southern France to 300 km south of the coast. We assume a broken plate boundary along the African coast. (Flexure of the European coast is essentially the same for a broken plate as it would be for a plate unbroken at the African margin.) The calculation was first done with a two-dimensional seawater load ('Water' in Fig. 1) and was repeated for several flexural rigidity values.

The removal of the load of only the present sea water gives us a flexural response representing the uplift caused by the last desiccation before the final infilling. Another calculation was made for removal of water and the 1.6-km-thick evaporites ('Salt' in Fig. 1). This unloading corresponds to the first Messinian dehydration of the Mediterranean, before its depth had been reduced by deposition of the thick evaporites. Finally, the flexure profiles were superimposed on the present-day topography.

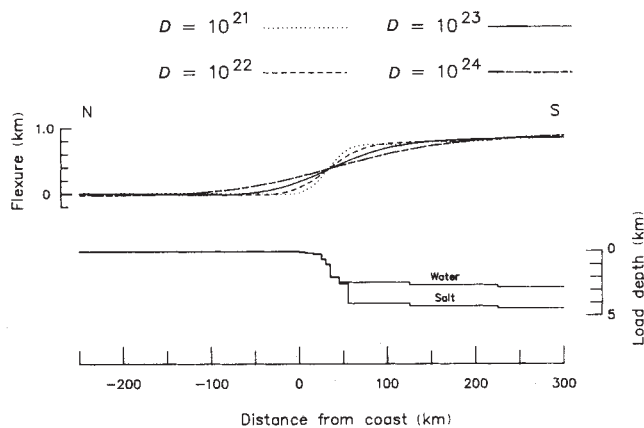


Fig. 1 Uplift of the Mediterranean Basin and 250 km inland on the French coast due to unloading of 2.6 km of water (density $1,000 \text{ kg m}^{-3}$) for a range of values of flexural rigidity D (in N m). Load depth profiles are shown at the bottom. The flexure profiles are calculated for removal of present water load only. Topographic relief is not considered.

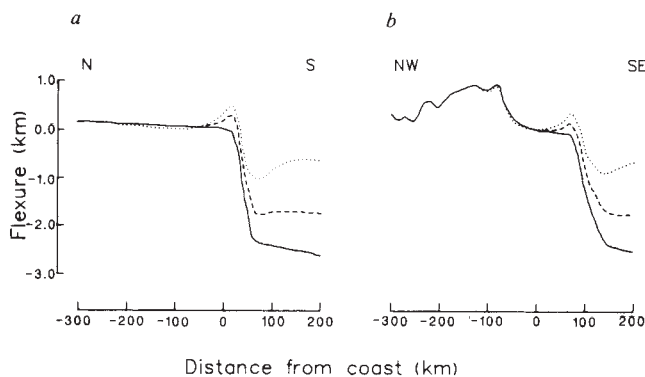


Fig. 2 Flexural effects combined with present topography, giving a regionally compensated palaeotopography profile from 200 km offshore to 300 km onshore, assuming $D = 10^{23} \text{ N m}$. Solid line, present profile; broken line, latest Messinian palaeotopography, calculated by removal of water load; dotted line, early Messinian palaeotopography, calculated by removing both water and salt loads (density $2,200 \text{ kg m}^{-3}$). *a*, N7E through Rhone Valley, slight uphill gradient. *b*, N38W through Cape d'Agde, moderate uphill gradient.

Time-dependent viscous response of the upper mantle should not seriously affect our results. For the size of load and range of flexural rigidities considered here, and for reasonable mantle viscosities, the viscous relaxation time will be $< 10^4 \text{ yr}$ (ref. 6), while the time span of Messinian dehydration is probably $\sim 10^6 \text{ yr}$, with episodes of partial refilling^{2,7,8}.

Figure 1 shows the calculated deflection patterns for the seawater load; the effect of superimposing topography on the calculated uplift from seawater and evaporite load is illustrated in Fig. 2. In calculating these palaeotopographic profiles, we work back from the present and show where the present-day topographic surface would have been deformed by the flexural uplift. We have not estimated the pre-Messinian palaeotopography.

For all flexural rigidity values, the maximum uplift at the centre of the ocean basin is 0.85 km for removal of the seawater load, and 1.9 km for removal of the sea water and evaporites combined. Watts and Ryan⁹, employing a thermal subsidence model in this area, found the best value for the flexural rigidity D to be 10^{23} N m (this is used for all profiles in Fig. 2). Figure 2a through the Rhone valley (elevation gradient 0.5 m km^{-1})

has an elevation at the shelf edge greater than that 50–300 km inland. The magnitude of the elevation difference and the location of the axis of lowest topography depends on the flexural rigidity. For $D = 10^{23} \text{ N m}$ the shelf edge would have been raised 0.18 km above present sea level at the last desiccation (present seawater load removed) and 0.45 km at the first desiccation (seawater and evaporite load removed). The axis of lowest elevation would be 125 km inland from the present coastline for both models. The slight moat immediately off the shelf break is caused by the interaction of the sudden elevation drop with the increase in flexural uplift near the edge of the load. This moat is best developed for the initial uplift (dotted lines in Fig. 2).

Shoreline bulges with inland flexural depressions lead us to consider drainage reversals towards inland basins. There is, however, a problem with applying this to some of the major rivers of the Mediterranean, especially the Rhone. The presence of palaeogorges buried 1 km below present river valleys^{1,2,10} indicates that seaward flow probably persisted throughout the desiccation period. There is still the possibility that further upstream there was a drainage divide, or that during early stages of uplift there was a drainage reversal, but headward cutting (in this case, northward from the Mediterranean) eventually carved through the drainage divide, and cutting of the gorge at the coast was a later event.

The necessary condition for a reversal of drainage is that the change of gradient due to the uplift be faster than the downcutting rate of rivers. Even if large rivers, by virtue of their larger discharge and therefore increased excavating potential, managed to continue flowing in the same direction throughout the Messinian, this model predicts that smaller rivers, with their inherently weaker cutting potential would probably have reversed their drainage direction. Figure 2b demonstrates that a steeper original topographic gradient brings the on-land flexural basin closer to the palaeo-shore and makes it narrower. In addition to the formation of these shore-parallel moats, drainage patterns on the plateaus and in mountains were probably also altered by flexural warping.

This discussion is quantitatively applicable only to the north-western Mediterranean, and the accuracy there is limited by the assumption of two-dimensionality. The same sort of model should apply to the southeastern Mediterranean, which also has thick evaporites³, although there the broken plate boundaries are more complex and the basin is narrower.

Topographic effects of flexural upwarping due to evaporation of the Mediterranean waters depend on the original topography and flexural rigidity. For southern France, southern Spain, northern Egypt and northern Libya, in areas with topography originally flat enough to have a shoreline bulge, major rivers probably excavated through the bulge as it formed, while low-discharge rivers probably changed their course toward inland drainage basins. Uplift attending the first desiccation of the Mediterranean, before thick evaporite deposition, should have had about twice the amplitude of the uplift accompanying the last dehydration. Messinian palaeocurrent indicators, location of inland basins and better information on the depth, extent and character of palaeogorges on such rivers as the Rhone, as well as smaller rivers, would test these predictions.

Received 2 October 1985; accepted 1 May 1986.

1. Hsu, K. J. in *Messinian Events in the Mediterranean*. *Geodynamics Scientific Report No. 7* (ed. Drooger, C. W.) 60–67 (North-Holland, Amsterdam, 1973).
2. Hsu, K. J., Cita, M. B. & Ryan, W. B. F. *Init. Rep. DSDP 13*, 1203–1231 (1973).
3. Montadert, L., Letouzey, J. & Mauffret, A. *Init. Rep. DSDP 42*, 1037–1050 (1978).
4. Hetenyi, M. *Beams on Elastic Foundation* (University of Michigan Press, Ann Arbor, 1946).
5. Jordan, T. E. *Bull. Am. Ass. Petrol. Geol.* **65**, 12, 2506–2520 (1981).
6. Cathles, L. M. *The Viscosity of the Earth's Mantle* (Princeton University Press, 1975).
7. Hsu, K. J. *et al. Nature* **267**, 399–403 (1977).
8. Cita, M. B. *Am. Geophys. Un. Geodyn. Ser.* **7**, 113–140 (1982).
9. Watts, A. B. & Ryan, W. B. *Tectonophysics* **36**, 25–44 (1976).
10. Clauzon, G. *Bull. Soc. géol. Fr.* **24**, 567–610 (1982).

Thermoluminescence dating of Le Moustier (Dordogne, France)

H. Valladas*, J. M. Geneste†, J. L. Joron‡
& J. P. Chadelles§

* Centre des Faibles Radioactivités, Laboratoire Mixte CNRS-CEA, Avenue de la Terrasse, 91190 Gif-sur-Yvette, France
† UA 133 du CNRS; Direction des Antiquités Préhistoriques d'Aquitaine, 26-28 Place Gambetta, 33074 Bordeaux Cedex, France
‡ Groupe des Sciences de la Terre, Laboratoire Pierre Sûe, CEN, Saclay, 91191 Gif-sur-Yvette, France
§ Direction des Antiquités Préhistoriques d'Aquitaine, 26-28 Place Gambetta, 33074 Bordeaux Cedex, France

The chronology of the Neanderthal cultures, commonly called Mousterian, which flourished between ~100,000 and 40,000 yr ago, has not been well established. The lower rock-shelter of Le Moustier provides one of the most important stratigraphic sequences, incorporating ten main Mousterian layers originally identified by Peyrony¹ and two early Upper Palaeolithic layers, the stratigraphy of which is mainly defined in terms of the regional climatic chronostratigraphy derived from the alpine glacial sequence¹⁻⁴. A correlation has recently been proposed between the chronostratigraphy of Dordogne and the past 125,000 yr of the well-established stratigraphy based on oceanic oxygen isotope ratios^{5,6}, but the lack of absolute datings of Mousterian settlements makes this correlation uncertain between 100,000 and 40,000 yr BP. We now report dates obtained by thermoluminescence measurements of 34 specimens of burnt flint recovered from the upper 3 m of the Mousterian deposits and from the superimposed Upper Palaeolithic layer. The dates, which range from 56,000 to 40,000 yr BP, allow us to correlate the cold sediment deposits of Le Moustier with cold intervals of the oxygen isotopic record of the Mediterranean Sea during isotopic stage 3.

Of the two major rock-shelters discovered at Le Moustier at the end of the last century, the upper or classic shelter was studied by Lartet and Christy (1863), whereupon Mortillet named the middle Palaeolithic industries Mousterian. The upper shelter was excavated at the beginning of the century^{1,7}. In the other shelter, which is located ~13.5 m below, several Mousterian traditions have been recognized^{1,2}, including typical Mousterian (layers B and J), Denticulate (layers F and I), and Acheulian tradition type A (layer G) and type B (layer H). The stratigraphic sequence, which is ~4 m thick, reveals two types of deposits: (1) the lower group of strata (layers A to F) is made up of waterlain sediments deposited by the floods of the Vézère River; (2) the upper group (layers G to L, Fig. 1a) consists of cryoclastic sediments in which the proportion of anthropological remains can be substantial (especially in layer H).

Each of the 12 main layers has been subdivided on the basis of sedimentation and pollen count data^{4,8}, and the whole sequence has been assigned to the early part of the last glacial period (ancient Würm in the chronostratigraphy of Dordogne^{5,6}). Layer K, overlying the most recent Mousterian layer, contains a mixture of Lower Perigordian (Chatelperronian) and Mousterian tool assemblages. The state of preservation of the Mousterian component in this assemblage suggests possible contamination from below. Layer K has been assigned to the beginning of the recent Würm²⁻⁶, ~35,000 yr ago.

Burnt flints have been widely used for thermoluminescence (TL) dating⁹⁻¹⁴; they are common near prehistoric hearths, where some will have been heated to a temperature high enough to erase the thermoluminescence acquired by the minerals during their geological history. Good agreement among the TL dates obtained¹⁵ with four quartz and three flint fragments for the upper Perigordian layer (Gravettian) of La Vigne Brun (Villerest,

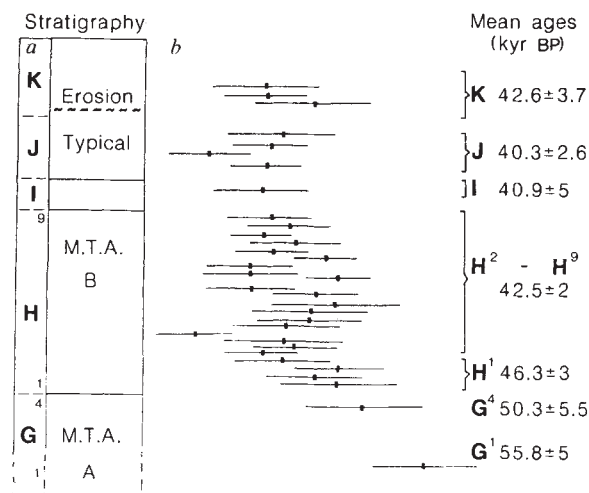


Fig. 1 a, The top Mousterian layers and the Upper Palaeolithic layers of the lower rock-shelter at Le Moustier^{1,4}. The thickness of this sequence is ~3 m. M.T.A., Mousterian of Acheulian tradition, type A (A) and type B (B). b, Horizontal bars representing the TL ages of the burnt flints from Le Moustier as a function of depth. The weighted ages calculated for the respective layers are listed at the right.

Loire, France) suggests that the TL of flint does not fade significantly.

The burnt flints were collected from a small area (30 × 40 cm) excavated in May 1982 in the north-east corner of the 'witness' sediments (level G to K). Approximately 80 flints of suitable size, showing signs of having been heated, were selected for TL examination. Of these, only 34 were sufficiently well burnt and had satisfactory TL characteristics. Three of these were from the middle part of layer K, 4 from layers J5 to J1, 1 from layer I3, 24 from layers H9 to H1 and 2 from layers G4 and G1. The bottom 1 m of Mousterian deposits (F to A) has not yet yielded burnt flints.

The methods used to treat the samples and to measure their palaeodose and annual dose rate have been described in detail elsewhere¹⁵. The environmental γ dose rate was measured using 28 calcium sulphate dosimeters in copper capsules. These were inserted at the ends of 30-cm-deep horizontal auger holes, which were distributed at ~10-cm intervals from top to bottom, and left buried for 1 yr. As the archaeological layers are also ~10 cm thick, it was impossible to specify the exact location of each sample within a specific layer. The external dose rate received by the flints of a given layer was obtained by taking the average dose measured in the layer itself and in the two adjoining layers, and ranges from 0.33 to 0.54 mGy yr⁻¹ except in layer G. The uranium and thorium contents in the flints were measured by epithermal neutron activation inside a cadmium container^{16,17}, and the potassium content by neutron activation at the Institut Pierre Sûe (CEN, Saclay). The internal dose rate usually represents ~50% of the total dose rate, which ranges from 0.67 to 0.90 mGy yr⁻¹. The results are reported in Table 1.

The age of each flint and its statistical error are presented as a function of depth in Fig. 1b. The ages (weighted mean) derived for the respective layers range from ~56,000 to 40,000 yr BP (Table 2). Levels J and H, for which many burnt flints were available, have been dated with better precision (5-6%) than the other levels; for the latter, <3 samples were available and the precision of the results is only ~10%.

The time span represented by the layers from which the dated flints were recovered (G to K) is estimated at 15,000 ± 5,000 yr. The paucity of data makes it impossible to determine whether there were any occupation gaps during the ~10,000 yr separating layers G1 from H1. On the other hand, virtually continuous

Table 1 Thermoluminescence results and radioactivity data for the flints from Le Moustier

Layer	Sample	U* (p.p.m.)	Th* (p.p.m.)	K* (p.p.m.)	$D_{\alpha}^{\dagger}$	$D_{\beta}^{\ddagger}$	External dose§	Dose rate	Palaeodose (10^{-2} Gy)	Age (kyr BP)
					$(10^{-5} \text{ Gy yr}^{-1})$					
K	3	0.823	0.473	334	22.2 ± 5.1	16.1 ± 0.9	44 ± 6	83 ± 8	$3,423 \pm 218$	41.3 ± 4.7
K	6	0.94	0.487	458	22.4 ± 2.7	18.8 ± 1.1	45 ± 6	87 ± 6.7	$3,595 \pm 185$	41.5 ± 3.8
K	7	0.915	0.256	296	24.7 ± 3.9	16.5 ± 0.9	44.8 ± 6	86.5 ± 7	$3,964 \pm 322$	45.8 ± 5.2
J5	24	0.687	0.352	300	19.2 ± 2	13.5 ± 0.8	42.7 ± 7	76 ± 7	$3,282 \pm 220$	43.2 ± 5
J2	52	0.801	0.455	394	23 ± 2.2	16.2 ± 0.9	40 ± 2	80 ± 3	$3,334 \pm 252$	41.7 ± 3.5
J2	53	0.647	0.657	450	19.2 ± 4.2	15 ± 0.9	40.5 ± 2	75.2 ± 5	$2,715 \pm 210$	36.1 ± 3.7
J1	71	0.668	0.28	300	21.2 ± 3.8	13 ± 0.7	38.3 ± 3	72.8 ± 5	$3,029 \pm 161$	41.6 ± 3.6
I3	120	0.772	0.29	100	19.7 ± 2.7	12.9 ± 0.7	36.2 ± 5	69.3 ± 5.8	$2,834 \pm 194$	40.9 ± 4.4
H9	141	0.851	0.54	350	21.6 ± 2	16.8 ± 1	32 ± 3.5	71.4 ± 4	$2,989 \pm 223$	41.9 ± 3.9
H9	142	0.719	0.62	430	17 ± 2.2	15.8 ± 0.9	32.6 ± 3.5	66 ± 4.2	$2,863 \pm 286$	43.4 ± 5.1
H9	150	0.907	0.42	360	23 ± 2.2	17.4 ± 1	33 ± 3.5	74 ± 4.2	$3,040 \pm 162$	41.1 ± 3.2
H7D	200	0.844	0.31	340	21.9 ± 2.2	16 ± 0.9	30.8 ± 3	69.7 ± 3.8	$3,088 \pm 265$	44.3 ± 4.5
H7C	220	1.27	0.4	370	37.3 ± 4.3	22.7 ± 1.3	31.4 ± 3	92.5 ± 5.4	$3,907 \pm 227$	42.2 ± 3.5
H7B	242	0.715	0.63	500	19.5 ± 1.4	16.3 ± 0.9	30.4 ± 3	67.5 ± 3.5	$3,167 \pm 127$	46.9 ± 3
H4	302	1.04	0.354	360	34.3 ± 4	19.1 ± 1.1	32.9 ± 3	86.9 ± 5.1	$3,471 \pm 256$	39.9 ± 3.8
H2E	360	0.93	0.3	300	29.5 ± 6	16.9 ± 1	35.1 ± 5	82.4 ± 8	$3,298 \pm 196$	40 ± 4.4
H2E	362	1.11	0.4	421	26.5 ± 2.6	20.8 ± 1.2	35.4 ± 5	83.5 ± 5.7	$3,301 \pm 245$	39.5 ± 4
H2E	363	0.462	0.944	630	16 ± 4.1	14.6 ± 0.8	35.5 ± 5	66.7 ± 6.5	$3,209 \pm 243$	48.1 ± 5.8
H2B	401	0.9	0.34	100	23 ± 2.6	14.9 ± 0.8	35.3 ± 5	74 ± 5.7	$2,979 \pm 183$	40.3 ± 4
H2B	402	0.89	0.35	333	26 ± 2.2	16.7 ± 1	34.6 ± 5	78.3 ± 5.5	$3,594 \pm 183$	45.9 ± 4
H2B	407	0.53	0.35	160	15.2 ± 1.5	10 ± 0.6	35 ± 5	60.8 ± 5.2	$2,908 \pm 270$	47.8 ± 6
H2B	421	0.4	0.378	170	9.9 ± 0.9	8.3 ± 0.4	35.3 ± 5	53.9 ± 5.1	$2,458 \pm 191$	45.5 ± 5.6
H2B	423	1.07	0.454	440	29 ± 3.3	20.6 ± 1.2	34.6 ± 5	85.4 ± 6.1	$3,861 \pm 337$	45.2 ± 5.1
H2B	428	0.56	0.243	400	14.1 ± 2.6	12.1 ± 0.7	35.1 ± 5	62 ± 5.7	$2,693 \pm 173$	43.4 ± 4.9
H2A	443	1.32	0.45	440	40 ± 5.3	24.1 ± 1.4	34.3 ± 5	100 ± 7.4	$3,535 \pm 249$	35.3 ± 3.6
H2A	464	0.64	0.285	280	23.5 ± 3.8	12.4 ± 0.7	35 ± 5	71.6 ± 6.3	$3,058 \pm 305$	42.7 ± 5.7
H2A	480	1.25	0.525	380	29.9 ± 2.7	22.8 ± 1.3	35 ± 5	88.9 ± 5.8	$3,884 \pm 266$	43.7 ± 4.2
H2A	482	1.01	0.462	400	30.8 ± 4.4	19.3 ± 1.1	35.1 ± 5	86.2 ± 6.7	$3,551 \pm 134$	41.2 ± 3.6
H1	500	0.86	0.32	240	28 ± 2.5	15.4 ± 0.9	39 ± 5	83.1 ± 5.7	$3,597 \pm 277$	43.2 ± 4.5
H1	501	1.05	0.39	390	30.6 ± 3	19.6 ± 1.1	39.3 ± 5	90.3 ± 5.9	$4,361 \pm 274$	48.3 ± 4.4
H1	502	1.12	0.44	340	32.5 ± 3	20.4 ± 1.2	38.7 ± 5	92.7 ± 5.9	$4,274 \pm 346$	46.1 ± 4.8
H1	503	0.844	0.336	220	21.3 ± 5	15.1 ± 0.9	39.1 ± 5	76.1 ± 7.1	$3,636 \pm 247$	47.7 ± 5.6
G4	570	0.594	0.76	400	16.9 ± 2	14.1 ± 0.8	47 ± 5	78.5 ± 5.5	$3,952 \pm 277$	50.5 ± 5.5
G1	571	1.3	0.3	270	33.9 ± 4.4	22 ± 1.3	84 ± 8	141 ± 9	$7,864 \pm 342$	55.8 ± 5

* The values of U, Th and K each have an error of $\pm 6\%$.

† The β equivalent dose rate per unit of α flux was measured as in ref. 27 and the β equivalent annual dose of flints was calculated using specific fluxes deduced from ref. 28.

‡ The β dose in flints was calculated using the specific values given in ref. 29. The γ internal dose is $<0.01 \text{ mGy yr}^{-1}$.

§ The external dose rate includes a cosmic contribution of $5\text{--}6 \times 10^{-5} \text{ Gy yr}^{-1}$. It has been calculated using the data given in refs 30 and 31; the estimated average thickness of overburden is 9–10 m.

occupation is suggested by the many layers falling within the shorter time period separating layers H1 and K ($6,000 \pm 3,000 \text{ yr}$). Note that, although the ages obtained for the highest Mousterian layer J ($40,300 \pm 2,600 \text{ yr}$) and the overlying layer K ($42,600 \pm 3,200 \text{ yr}$) are not in accord with the stratigraphy, the magnitude of the experimental errors makes the inversion more

apparent than real. Also, as noted earlier, layer K contained a mixture of Lower Perigordian (Chatelperronian) and Mousterian tools. Thus, it may have been Mousterian flints which were responsible for the date obtained for layer K.

These dates allow us to correlate the stratigraphy of Le Moustier^{4,5} with the isotopic record and chronology of deep-sea sediments. The ages obtained for levels G to K place their deposition during isotopic stage 3 (ref. 18). A detailed chronology of this stage has recently been proposed by correlating five ash layers contained in cores collected in the central Tyrrhenian Sea with terrestrial volcanic deposits dated by the ^{14}C and/or the K–Ar method¹⁹. The temperature variations of the Tyrrhenian Sea are much larger than those of the open ocean; consequently, the Tyrrhenian Sea sediments provide an oxygen isotopic signal which is sensitive to both the water temperature and changes in the $\delta^{18}\text{O}$ content of the water^{20–22}. Nevertheless, the Tyrrhenian Sea, like the entire Mediterranean Sea, has yielded isotopic records showing features similar to those in the generally accepted oxygen isotope curve²³; indeed, the classical isotopic stages¹⁸ are easily recognized and within stage 3, a strong climatic variability with a succession of cold and warmer isotopic events has been described¹⁹.

The Tyrrhenian record shows the same climatic trends as those deduced from the analysis of continental pollen in

Table 2 Weighted mean ages of levels in which flints were found

Layers	Mean ages (yr BP)
K	$42,600 \pm 3,200$
J	$40,300 \pm 2,600$
I	$40,900 \pm 5,000$
H ₉ to H ₂	$42,500 \pm 2,000$
H ₁	$46,300 \pm 3,000$
G ₄	$50,300 \pm 5,500$
G ₁	$55,800 \pm 5,000$

The errors (statistical plus systematic) at 68% confidence level have been assessed as in ref. 32. A systematic error of $\pm 7.5\%$ has been assumed for the external dose, to take into account possible past variations of the water content of the archaeological sediments (M. J. Aitken, personal communication).

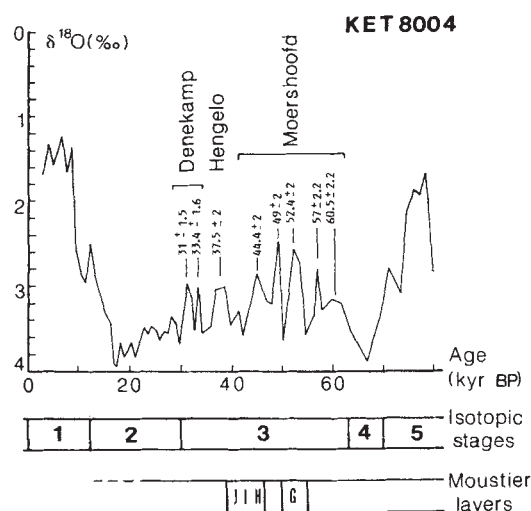


Fig. 2 The dated Moustier layers are shown below the $\delta^{18}\text{O}(\text{‰})$ variations measured in the core KET 8004 collected in the Tyrrhenian Sea. Estimates of the age of the oxygen isotopic events (light $\delta^{18}\text{O}$ peaks) have been calculated assuming a constant sedimentation rate between two well-dated ash layers¹⁹

northern Europe, as well as in France and eastern Macedonia²⁴⁻²⁶; the well-dated oxygen isotope peaks (warm events) which occur between 60,000 and 30,000 yr BP (ref. 19) correlate with the continental interstadials dated by the ^{14}C method²⁶. This record provides a fairly good comparison framework for the prehistoric stratigraphy, as it has been obtained in a climatic zone close to southern France.

The dated layers of Le Moustier are shown below the isotopic record of the Tyrrhenian Sea (Fig. 2). The cryoclastic sediments of layers G to K have yielded low pollen counts (<10%), which suggests that these layers were deposited during cold periods^{4,8}. The TL ages obtained for the above-mentioned layers suggest that their deposition occurred during the cold periods which developed between 56,000 and 40,000 yr BP on the isotopic record.

We therefore conclude that TL dating of burnt flints should make it possible to obtain close correlation between the floating chronology of the Mousterian epoch based on stratigraphic evidence and the more absolute chronology derived from the isotopic record of sea cores.

The field research at Le Moustier has been supported by the L.S.B. Leakey foundation, Pasadena, California (grant for the Combe-Saunière Dating Project, 1981). We thank G. Valladas for his important contribution to this work and J. L. Reyss for radioactivity analyses. CFR contribution 760.

Received 5 March; accepted 30 April 1986.

- Peyrony, D. *Rev. Anthropol.* 1-3; 48-76; 4, 6 (1930).
- Bordes, F. *Libret Guide de l'excursion A4, Landes-Perigord*, 38-87 (VIII Congrès INQUA, Paris, 1969).
- Laville, H. & Rigaud, J. P. C. *r. hebdomadaire Séanc. acad. Sci., Paris D* 276 3097-3100 (1973).
- Laville, H. in *Etudes Quaternaires Mémoires* (Marseille), 4 (1975).
- Laville, H. *et al. Paleoclimat*, 34, 219-241 (1983).
- Laville, H., Raynal, J. P. & Texier, P. J. in *Bull. Soc. préh. franç.* 81, 8-11 (1984).
- Bourlon, M. *L'homme préhistorique* 3, 7, 14 (1905).
- Paquereau, M. M. *Quaternaria* 18, 85-93 (1974-1975).
- Gosku, H. Y., Fremlin, J. H., Irwin, H. T. & Fryxell, R. *Science* 183, 651-654 (1974).
- Wintle, A. G. & Aitken, M. J. *Archaeometry* 19, 111-130 (1977).
- Valladas, H. *PACT* 2, 180-184 (1978).
- Huxtable, J. & Jacobi, R. M. *Archaeometry* 24, 164-169 (1982).
- Bowman, S. G. E. & Sieveking, G. de G. *PACT* 9, 353-361 (1982).
- Aitken, M. J. *Thermoluminescence Dating* (Academic, New York, 1985).
- Valladas, H. thesis, Paris VI and Museum Nat. Hist. Nat. (1985).
- Chayla, B., Jaffrezic, H. & Joron, J. L. C. *r. hebdomadaire Séanc. Acad. Sci., Paris*, 277, 273-275, (1973).
- Joron, J. L. thesis, Orsay (1974).
- Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* 3, 39-55 (1973).
- Paterne, M., Guichard, F., Labeyrie, J., Gillot, P. Y. & Duplessy, J. C. *Mar. Geol.* (in the press).
- Emiliani, C. *Quaternaria* 2, 87 (1955).
- Vergnaud-Grazzini, C. *Science* 190, 272-274 (1975).

- Thumell, R., Federman, A., Sparks, S. & Williams, D. *Quat. Res.* 12, 241-253 (1979).
- Ryan, W. B. F. in *The Mediterranean Sea*, (Stanley, D. J. ed.) 149-170 (Dorden, Hutchinson and Ross, 1972).
- Van Der Hammen, T., Wijmstra, T. A. & Zagwija, W. H. in *The Late Cenozoic Glacial Ages* (Turekian, K. K. ed) 391-424 (Yale University Press, Cambridge, 1971).
- Woillard, G. *Quat. Res.* 9, 1-21 (1978).
- Woillard, G. & Mook, W. G. *Science* 215, 159-161 (1982).
- Valladas, H. & Valladas, G. *PACT* 6, 171-178 (1982).
- Aitken, M. J. & Bowman, S. G. E. *Archaeometry* 17, 132-138 (1975).
- Bell, W. T. *Archaeometry* 21 243-245 (1979).
- Kvasnicka, J. *Hlth Phys.* 36, 521-524 (1979).
- Prescott, J. R. & Stephan, L. G. *PACT* 6, 17-25 (1982).
- Aitken, M. J. *Archaeometry* 18, 233-238 (1976).

A putative ancestor for the swordfish-like ichthyosaur *Eurhinosaurus*

C. McGowan

Department of Vertebrate Palaeontology, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario, Canada M5S 2C6, and Department of Zoology, University of Toronto

Ichthyosaurs, extinct marine reptiles superficially similar to sharks and dolphins, flourished throughout most of the Mesozoic, but are best represented in the early Jurassic¹⁻⁵. Hundreds of skeletons are known, principally from the Lower Liassic (Hettangian, Sinemurian and Lower Pleinsbachian) of south-west England and the Upper Liassic (Toarcian) of southern Germany⁶⁻⁹. Among the most specialized is *Eurhinosaurus*, unique for its long rostrum and shortened mandible, which confers a striking resemblance to the modern swordfish, *Xiphias*^{1,5,10}. Known only from the Upper Liassic of Germany, this monotypic genus, like most fossil species, appears suddenly, without antecedents¹¹⁻¹⁴. The recent discovery of a geologically older (Lower Liassic) ichthyosaur with a partially extended rostrum is therefore of considerable interest. As described here, this specimen, which is too large to be part of a growth series of *Eurhinosaurus*, appears to be ancestral, and could have given rise to *Eurhinosaurus* through changes in growth rates of the rostrum and mandible.

The incomplete skeleton of this new ichthyosaur was collected in Somerset in 1984 by D. Costin. It is described below as representing a new genus.

Class Reptilia
Order Ichthyosauria
(Family designation awaits revision)
Excalibosaurus gen. nov.

Etymology: Excalibur, the mythological sword of Arthurian legend; *saurus*, Greek (masculine) for lizard.

Diagnosis: Mandible shorter than skull, but exceeding 60% of skull length. Snout extends well beyond anterior tip of mandible, but length of snout (distance between tip of snout and anterior margin of orbit¹⁵) not greatly exceeding length of mandible.

Type species: *Excalibosaurus costini* gen. et sp. nov. (Fig. 1).

Etymology: In honour of D. Costin, through whose diligence and generosity the material was collected and safely deposited in a museum.

Diagnosis: As for genus.

Holotype: Cc 881, a complete skull, with associated forefin, pectoral girdle, vertebrae and ribs, deposited in The City of Bristol Museum and Art Gallery, UK.

Locality and horizon: The foreshore near Lillstock, Somerset, UK (map ref. ST 196 463). Lower Liassic (Sinemurian), horizon provisionally given as the Bucklandi Zone (M. L. K. Curtis, personal communication).

Description: The elongated snout, which projects well beyond the tip of the mandible, distinguishes this from all other ichthyosaurs, except *Eurhinosaurus huenei*, where the disparity

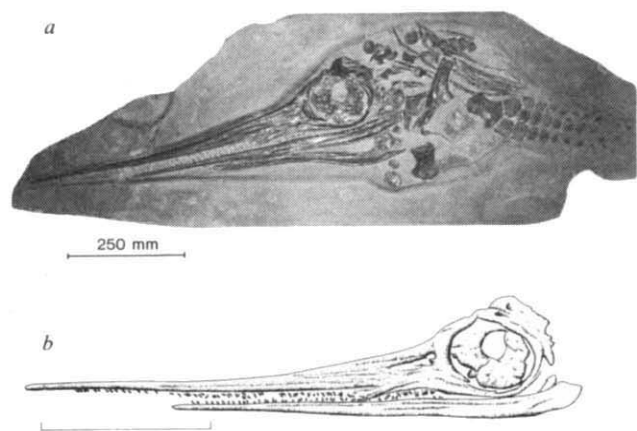


Fig. 1 *Excalibosaurus costini* gen. et sp. nov. Holotype (Cc881, City of Bristol Museum and Art Gallery). Lower Liassic (Sinemurian) of Lillstock, Somerset. *a*, Complete specimen; *b*, diagram of the skull. Scale bar, 250 mm.

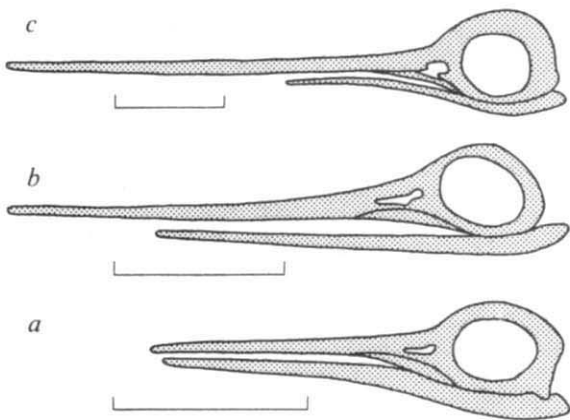


Fig. 2 Ichthyosaur skulls drawn to the same mandibular lengths (*a*, *b*), and to the same skull lengths (*b*, *c*). *a*, *Ichthyosaurus tenuirostris* (based on OUM J10305, Oxford University Museum, and IGS 51236, Institute of Geological Sciences, London) from Lower Liassic (Hettangian–Sinemurian), Street, Somerset. *b*, *Excalibosaurus costini* gen. et sp. nov., holotype, from Lower Liassic (Sinemurian), Lillstock, Somerset. The apparent difference in shape between *a* and *b* in the vicinity of the jaw joint is an artefact of preservation. *c*, *Eurhinosaurus huenei* (based on MNHP 1946–20, Museum National d'Histoire Naturelle, Paris, and FSF 4155, Forschungsinstitut Senckenberg, Frankfurt) from Upper Liassic (Toarcian), Holzmaden. The transition from *a* to *c*, via *b*, could have been effected through an elongation of the rostrum, followed by a reduction in the mandible. Scale bar, 250 mm.

between snout and mandible is even greater. The snout ratio¹⁵ (ratio of the distance between the tip of the snout and the anterior margin of the orbit, to the length of the mandible) is 1.05 (see Table 1). Although the skull is smaller (785 mm long) than that of *E. huenei*⁵ (observed range 1,090–1,550 mm), the mandible (600 mm long) falls within the observed range for that species (523–792 mm) so that the new material cannot be discounted as an immature specimen of *E. huenei*. The possibility that the long snout is an artefact is rejected because the mandible and snout both taper naturally and gently throughout their lengths, and the mandible has not been displaced caudally to any significant extent. There is an oblique crack in the matrix at the level of the tip of the mandible, but no obvious breaks in continuity in the rostrum, which is so much deeper than the mandible that it is inconceivable that it could have terminated here anyway. The possibility that the specimen is an abnormal individual of a known taxon cannot be verified.

Compared with most other Lower Liassic ichthyosaurs, the new species is relatively large, the skull length exceeding that of *Ichthyosaurus communis* and lying at the top end of the range for *I. tenuirostris*³. The new species is, however, smaller than most specimens referred to *Temnodontosaurus*². Its slender teeth extend along most of the free portion of the snout; *Eurhinosaurus huenei* similarly has a dentigerous 'sword'. Further preparation is needed, but the external naris appears bilobed, as in some

specimens of *E. huenei*⁵. A similar naris occurs in *I. tenuirostris*, which has other features in common with the new taxon. This similarity could be used to support a relationship between the three taxa, but it is not a constant feature, and a somewhat similar naris occurs in *Leptopterygius burgundiae*⁵. The naris of the new species appears to be relatively large. The prenarial¹⁵ ratio (ratio of the distance between the tip of the snout and the anterior margin of the external naris, to the length of the mandible) is relatively higher (0.89) than in *I. tenuirostris* (0.60). The anterior end of the maxilla has been displaced, but the premaxillary ratio¹⁵ (ratio of the distance between the tip of the snout and the anterior tip of the maxilla, to the length of the mandible) is estimated to be 0.86. The orbital ratio¹⁵ (ratio of the diameter of the orbit to the length of the mandible) is 0.20. The forefin has three digits, with some indication of a fourth; there are

Table 1 Cranial and mandibular ratios for selected ichthyosaurs								
Species	Cranial ratios				Mandibular ratios			
	Snout/ skull	Orbit/ skull	Premaxilla/ skull	Prenaris/ skull	Snout/ mandible	Orbit/ mandible	Premaxilla/ mandible	Prenaris/ mandible
<i>Excalibosaurus costini</i> gen. et sp. nov.	0.80	0.16	0.66	0.68	1.05	0.20	0.86	0.89
<i>Ichthyosaurus tenuirostris</i> (mean values)	0.75	0.23	0.54	0.62	0.72	0.21	0.52	0.60
Modified specimen of <i>I. tenuirostris</i> *	0.83	0.16	0.67	0.74	1.09	0.21	0.88	0.97
<i>Eurhinosaurus huenei</i> (mean values)	0.83	0.12	0.64	0.77	1.64	0.25	1.28	1.53
Modified specimen of <i>Excalibosaurus costini</i> †	0.80	0.16	0.66	0.68	1.58	0.31	1.29	1.33

Cranial ratios are measurements (such as the length of the snout) compared with the length of the skull, while in mandibular ratios they are compared with the length of the mandible.

* *I. tenuirostris* modified by increasing the snout length such that the ratio skull length/mandibular length is the same as in *E. costini*.

† *E. costini* modified by reducing the length of the mandible such that the ratio mandible length/skull length is the same as in *E. huenei*.

usually four in *I. tenuirostris* and up to five in *E. huenei*. The forefin is about half the length of the mandible, as in *I. tenuirostris*, whereas in *E. huenei* it is relatively longer, often exceeding the length of the mandible. There are 12 elements in the longest digit (commencing with the epipodials¹⁶), which is within the observed range for *I. tenuirostris*, but there are usually more in *E. huenei* (up to 23). Fin structure, however, is often variable within species, especially in *E. huenei*⁵.

Aside from its longer snout, the skull of *E. costini* is similar to that of *I. tenuirostris* (compare Fig. 2b, a); this results in their having similar orbital ratios, but their other mandibular ratios, and their cranial ratios, are disparate (Table 1). However, if the snout of *I. tenuirostris* is increased in length until the ratio of skull length/mandibular length corresponds to that of *E. costini*, their cranial and mandibular ratios converge (Table 1). *E. costini* could therefore be derived from *I. tenuirostris* by an elongation of the snout.

E. costini is cranially similar to *E. huenei*, except in having a relatively longer mandible (Fig. 2b, c); this is reflected in the similarity of their cranial ratios, and the dissimilarity of their mandibular ratios (Table 1). However, if the mandible of *E. costini* is reduced to 400 mm, giving the same value for the ratio mandible length/skull length as in *E. huenei* (0.51), its mandibular ratios converge with those of *E. huenei* (Table 1). *E. huenei* could therefore be derived from *E. costini* by mandibular reduction.

If *E. huenei* evolved from a long-snouted ancestor like *I. tenuirostris*, through an intermediary like *E. costini*, what mechanism might have been involved? A change in growth rates of the rostrum and mandible during ontogeny seems plausible, and some insight is gained by studying ontogeny in *Xiphias*. Larval swordfish have mandibles extending to the tip of the skull¹⁷⁻¹⁹ whereas in adults the mandible is only about one-third of the skull length. These profound changes are effected by relatively small differences in growth rates between the mandible and rostrum (C. McG., in preparation). A small increase in the growth rate of the rostrum in an ichthyosaur like *I. tenuirostris* could therefore give rise to a descendant like *E. costini* with an elongated rostrum. This, in turn, could give rise to *E. huenei* by a reduction in the growth rate of the mandible. Such minor modifications would presumably require relatively small genetic changes, but the morphological, and consequent biological changes would be profound.

I thank M. A. Taylor for bringing the new material to my attention and also thank him, S. Swansborough and M. L. K. Curtis for their help in Bristol. I thank J. Mulock for his drawings; P. Purves for typing; R. Johnson and J. Thomason for reading the manuscript; C. E. McGowan for nomenclatural inspiration; and J. Campsie for classical advice. This research was supported by the Natural Sciences and Engineering Research Council of Canada, grant A 9550.

The genetic control and consequences of kin recognition by the larvae of a colonial marine invertebrate

Richard K. Grosberg* & James F. Quinn†

* Department of Zoology and † Division of Environmental Studies, University of California, Davis, California 95616, USA

The evolution of altruism, cooperation and sociality should be favoured by mechanisms promoting interactions among relatives^{1,2}. In turn, the opportunity for such interactions should be enhanced where related individuals are spatially associated. The simplest explanation for association of kin invokes philopatric, or limited, dispersal³. Alternatively, kin recognition—which is known from a broad array of taxa^{4,5}—can produce similar associations. Neither the prevalence of kin recognition, nor aggregations of kin, are by themselves sufficient to demonstrate that kin recognition plays an important role in the production of nonrandom associations of relatives. To have such a role, kin recognition must promote or inhibit associations of kin beyond the effects of other processes, notably dispersal, that modify spatial patterns. Here we report the results of field experiments showing that sibling planktonic larvae of the sessile colonial ascidian *Botryllus schlosseri* settle in aggregations that are much stronger than expected from dispersal distance effects alone. Laboratory experiments indicate that larvae distinguish kin on the basis of shared alleles at a highly polymorphic histocompatibility locus known to regulate fusion between adult colonies. This kin recognition mechanism, along with limited dispersal of larvae, promotes co-settlement of histocompatible individuals. Consequently, the probability of fusion between adult colonies is far greater than that expected if larvae settled randomly.

Botryllus schlosseri colonies are founded after a sexually produced planktonic tadpole larva attaches to a firm substratum and metamorphoses. Repeated cycles of asexual multiplication produce a colony of morphologically and genetically identical zooids, connected by a blood vascular system. As in many colonial marine invertebrates, the larvae of several oviparous ascidians can metamorphose soon after release into the plankton⁶⁻¹⁰. In the case of *B. schlosseri*, our laboratory and field observations, as well as those of others¹¹, demonstrate that larvae can metamorphose—and many do—soon after escaping from their natal colony.

To determine the contribution of larval settlement behaviour to the spatial arrangement of sibling adult colonies, we examined the recruitment pattern of sibling colonies (<4 weeks old) that were founded by larvae, all derived from the same cross, and carrying a rare genetic marker. We identified the marker from an electrophoretic survey of 512 colonies living on the side of the Marine Biological Laboratory Supply Dock in the Eel Pond at Woods Hole, Massachusetts, USA. At the phosphoglucose isomerase locus (*PGI*), five electromorphs occurred, one of which (*PGI-fast*) was present in only two of the colonies surveyed. We cross-fertilized these two colonies, then assayed the *F*₁ colonies for *PGI-fast* homozygotes. One *F*₁ homozygote was selected as the source colony for marked larvae, and was positioned at the centre of a 1-m diameter circular asbestos-cement panel horizontally suspended 0.5 m below the dock. All zooids in a sexually mature *B. schlosseri* colony ovulate synchronously approximately weekly. Just before each ovulation, we returned the source colony to the laboratory, where we mated it with a sibling colony homozygous for the same marker allele, then repositioned the source colony on the panel. We left the panel in position for 4 weeks, then mapped and removed all *B. schlosseri* colonies that had recruited. Colonies were frozen at -80 °C, then analysed electrophoretically. We mapped the panel before any recruited colonies had reached sexual maturity; thus,

Received 6 January; accepted 1 May 1986.

1. Von Huene, F. Die Ichthyosaurier des Lias und ihre Zusammenhänge (Verlag von Gebrüder, Borntraeger, Berlin, 1922).
2. McGowan, C. *Life Sci. Contr. R. Ont. Mus.* **93**, 1-37 (1974a).
3. McGowan, C. *Life Sci. Contr. R. Ont. Mus.* **100**, 1-30 (1974b).
4. McGowan, C. *J. Paleont.* **52**, 1155-1162 (1978).
5. McGowan, C. *Palaeontographica* **166**, 93-135 (1979).
6. Hauff, B. *Palaeontographica* **64**, 1-42 (1921).
7. Hauff, B. *Das Holzmadenbuch* (F. Rau, Öhringen, 1953).
8. Arkell, W. J. *The Jurassic System in Great Britain* (Oxford University Press, 1933).
9. Palmer, C. P. *Newsl. Stratigr.* **2**, 45-54 (1972).
10. Jäger, G. *Nova Acta Akad. Caesar. Leop. Carol.* **25**, 937-967 (1856).
11. Darwin, C. *On the Origin of Species* (Murray, London, 1859).
12. Eldredge, N. & Gould, S. J. in *Models in Paleobiology* (ed. Schopf, T. J.) 82-115 (Freeman, Cooper, San Francisco, 1972).
13. Gould, S. J. *Paleobiology* **11**, 2-12 (1985).
14. Gingerich, P. D. *Paleobiology* **11**, 27-41 (1985).
15. McGowan, C. *Can. J. Earth Sci.* **13**, 668-683 (1976).
16. McGowan, C. *Life Sci. occ. Pap. R. Ont. Mus.* **20**, 1-8 (1972).
17. Arata, G. F. *Bull. mar. Sci. Gulf Caribb.* **4**, 183-243 (1954).
18. Yabe, H., Ueyanagi, S., Kikawa, S. & Watanabe, H. *Rep. Nankai reg. Fish Res. Lab.* **10**, 107-150 (1959).
19. Markle, G. E. *U.S. Dept. Commerce, NOAA tech. Rep. NMFS SSRF-675*, 252-260 (1974).

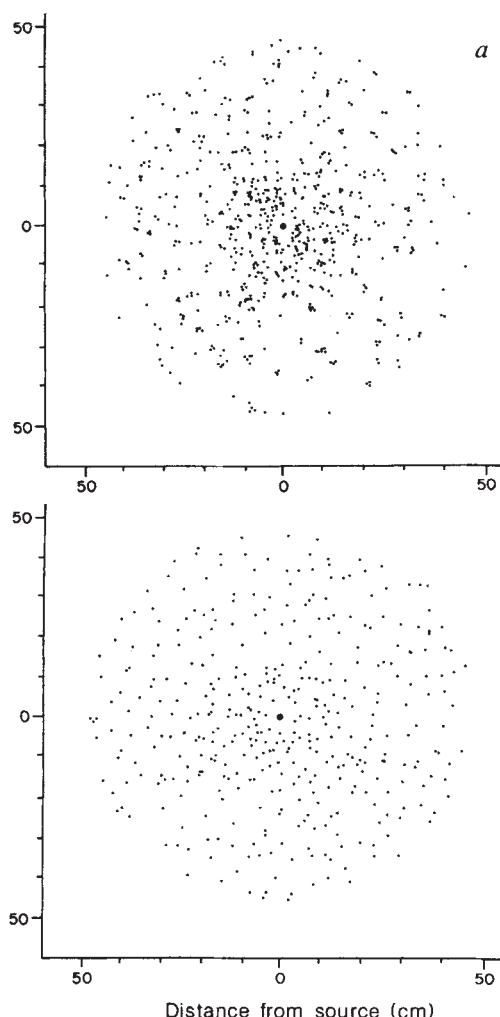


Fig. 1 Maps of the dispersion of *Botryllus schlosseri* (which settled on the undersurface of a 1-m circular asbestos-cement panel suspended below the Marine Biological Laboratory Supply Dock at Woods Hole, Massachusetts). Small dots depict the locations of colonies. The large dot marks the location of the source colony. *a*, Map of F_1 colonies (identified by the presence of the *PGI-fast* allele, see text) derived from source colony. *b*, Map of colonies not carrying the *PGI-fast* allele, thus, not derived from the source colony.

colonies homozygous for the *PGI* marker allele were assumed to be siblings originating from the source colony at the centre of the panel. Recruits not carrying the marker allele must have immigrated.

Figure 1*a* and *b* show respectively maps of marked and unmarked colonies. To test for spatial association among siblings, we computed the Clark and Evans¹² index of aggregation (R) based on nearest-neighbour distances between sibling colonies. R is the mean nearest-neighbour distance observed, divided by that expected if dispersion were random. Thus, randomly arranged individuals are expected to yield a value for R of 1.0. Values of R significantly <1 indicate aggregation, whereas values >1 indicate overdispersion. The sibling colonies were aggregated as tested by Clark and Evans' significance criteria ($R=0.769$; $P<0.001$). The unmarked immigrant colonies were significantly overdispersed ($R=1.227$; $P<0.001$). This overdispersion could result from larval behaviour, or could simply be an artefact of spacing due to colony growth¹³. In any case, the degree of aggregation of siblings differs even more from that of immigrants than it does from random (t -test; $P<0.001$).

Figure 1*a* shows that sibling colonies are clustered around the source colony, perhaps because *B. schlosseri* larvae can settle

upon release into the plankton. To separate any aggregation caused by sibling recognition from that due to limited dispersal, we constructed a null model based on a bootstrap analysis¹⁴ of the nearest-neighbour data. The locations of colonies were expressed in polar coordinates with the origin at the centre of the source colony: pseudosamples with the same distributions of distances and directions were chosen independently and randomly, with replacement, from the actual coordinates of sibling colonies. The randomized distance and direction coordinates are thus independent in the pseudosamples. This randomization procedure should preserve in the pseudosamples the overall aggregating effect of dispersal distance alone. However, smaller-scale aggregations due to interactions between immediate neighbours will be broken up by the randomization of the angular coordinates. For 500 randomizations, we calculated the value of R produced by the random model. In all cases, the values of R under the random model were substantially greater (mean = 0.897; s.d. = 0.023; range 0.824–0.964) than the observed $R=0.769$. This analysis demonstrates that limited dispersal explains only about half the deviation of mean nearest-neighbour distance from that expected if settlement were random.

The histocompatibility system known to control fusion and rejection between adult *Botryllus* colonies may provide a mechanism by which larvae recognize their siblings. Histocompatibility is controlled by a single mendelian locus such that two colonies sharing one or both alleles can fuse by their blood vascular systems. When colonies do not share an allele, rejection, accompanied by necrosis at the site of allogeneic contact, occurs^{15–17}. Fusion frequencies among pairs of *Botryllus* colonies collected in the field have been reported to range from 4 to 8% (refs 18, 19). In our studies on the Marine Biological Laboratory Supply Dock, of 500 pairs of colonies collected haphazardly along a 20-m transect, 22 pairs (4.2%) fused. According to the method of Curtis *et al.*²⁰, at equilibrium, approximately 100 equally frequent histocompatibility alleles would produce the fusion frequency observed in the Eel Pond population. Given such a high level of polymorphism at the histocompatibility locus, the likelihood that any two colonies will share a histocompatibility allele should depend primarily on their pedigree relatedness.

Because the settled colonies in the field experiments were all full siblings, shared alleles at loci other than the histocompatibility locus could have been used for sibling recognition. Accordingly, we designed laboratory experiments to test whether larval settlement location is better predicted by the histocompatibility type or the relatedness of resident individuals. With a breeding programme using colonies of known histocompatibility genotypes (determined by fusion assays), in a factorial design, we varied independently the effects of pedigree relatedness and histocompatibility genotype of larvae and residents on distance between settled larvae and previously attached residents. Thus, there were four experimental treatments: (1) the residents and larvae were inbred, and shared one histocompatibility allele; (2) the residents and larvae were inbred, but did not share a histocompatibility allele; (3) the residents were fourth-generation outbred descendants of the parents of the larvae, with one shared histocompatibility allele; and (4) the same as (3), but residents and larvae did not share a histocompatibility allele. In outline, the breeding programme was as follows: for treatments (1) and (2), homozygous resident colonies of known histocompatibility genotype (AA) were derived from crosses between heterozygous sibling colonies (AB). Unfusable homozygous colonies (BB) were isolated from the same cross. In treatment (1), introduced larvae (AB) were derived from a cross between the AA and BB full-sibs. In treatment (2), introduced larvae (BB) were derived from a $BB \times BB$ cross from the same sibship used to generate the larvae in treatment (1). Therefore, the introduced larvae in treatments (1) and (2) bear the same pedigree relatedness to the residents, but differ in the presence

Table 1 Two-way analysis of variance on the distance between settled larvae and the nearest resident colony as a function of resident/larval relatedness and histocompatibility types

Source	Sum of squares	Degrees of freedom	F value (P)
Relatedness	57.2	1	1.35 (>0.25)
Histocompatibility type	3,708.5	1	87.46 (<0.001)
Replicate	39.8	2	0.47 (>0.63)
Relatedness/histocompatibility	174.1	1	4.10 (0.04)
Error	9,964.6	235	

Table 2 Summary of the mean distances between settled larvae and the nearest resident colony according to experimental treatment

Treatment	Mean (mm)	n
Siblings, no shared alleles	14.70	60
Siblings, one shared allele	7.80	60
Outbred, one shared allele	5.15	60
Outbred, no shared alleles	13.92	60

n, Number of observations for each treatment.

of an A histocompatibility allele. For treatments (3) and (4), either AA or BB larvae were introduced into Petri dishes with outbred residents carrying an A allele. These residents were derived after sequential outcrossing with different colonies taken from the field, none of which carried either an A or B allele. After each generation, a colony carrying the A allele was identified (with a fusion assay) and used for subsequent matings. After four generations of outcrossing, single zooids from a heterozygous colony carrying the A histocompatibility allele, and an unknown second allele (differing from B), were used as residents.

For each treatment, we established three replicates, each with 20 resident colonies. The resident colonies were founded by attaching single zooids from a colony of known genotype to random positions on the submerged undersurfaces of the tops of 150-mm plastic Petri dishes. Twenty-four hours later, the dishes were filled with seawater, sealed and kept in the dark. We then introduced 20 larvae through a port in the side of each dish according to the above treatments. One day later, we measured the distance from each settled larva to the nearest resident.

The results of an analysis of variance show that the relatedness of residents and larvae has no discernible effect on nearest-neighbour distances (Table 1). In contrast, where larvae and residents shared a histocompatibility allele, the larvae settled significantly closer to the residents (Table 1). Indeed, the mean nearest-neighbour distance between larvae and residents sharing a histocompatibility allele is roughly half that observed when no allele is shared (Table 2). Table 1 shows a marginally significant interaction between the two main effects that is difficult to interpret. However, the minor contribution that this interaction makes to the overall sum of squares suggests that the interaction is biologically unimportant in these experiments.

Our experiments do not eliminate the possibility that larval kin recognition and histocompatibility are closely linked, rather than identical, genetic traits. However, because larval kin recognition increases the likelihood of adjacent colonies being fusible—an otherwise improbable event given the polymorphism of the histocompatibility locus—there are functional grounds for expecting the two processes to be controlled by a single locus.

The promotion of co-settlement of histocompatible colonies, coupled with the restriction of fusion to closely related

genotypes, indicate that colony fusion may be beneficial, but only among kin. In colonial organisms such as *Botryllus*, fusion may benefit one, or both, members of a chimaera in several ways. First, colony fusion immediately increases colony size. Survivorship is known to be size-dependent in several colonial ascidians (refs 21, 22 and our unpublished observations), sponges^{23,24}, cnidarians^{25–28} and ectopods^{29,30}; hence, fusion may reduce the likelihood of total mortality^{29,30}. Second, if the onset of reproduction is size-dependent—as it appears to be in several cnidarians^{28,31} and ascidians^{6,21}—then fusion may lower the age at first reproduction of both members of the chimaera³². In colonial organisms where growth rate is size-dependent^{33,34}, fusion may also increase the probability of subsequent survival and reproduction. Indeed, fusion among juvenile colonial organisms is well known^{32,35}. Third, a chimaera may have different physiological attributes from either component colony, potentially increasing the range of environmental tolerance³². Finally, fused botryllid ascidians (like many other clonal taxa^{36,37}) freely exchange cells that can differentiate into gametes³⁸, and the rate of differentiation may vary genetically³⁸. Under these conditions, somatic cell parasitism, in which one member of a chimaera increases its reproductive output at the expense of the other, can occur³². However, because the polymorphism and genetics of the *Botryllus* histocompatibility system limit fusion almost entirely to closely related individuals, the parasitic losses, in evolutionary terms, would be considerably less than if the fused colonies were unrelated^{1,39}.

Another potential consequence of close association of kin is increased inbreeding. Laboratory studies suggest that self-fertilization is deleterious in *B. schlosseri*⁴⁰ and, in some congeners, syngamy may not occur if sperm and egg carry the same histocompatibility allele^{41,42}. However, there is no evidence of inbreeding depression in the Eel Pond population of *Botryllus schlosseri*⁴³. Indeed, studies of other marine invertebrates in which closely related individuals are likely to cross-fertilize provide meagre evidence for inbreeding depression^{9,44}.

Botryllus schlosseri appears to be one of the first species where a functional—and perhaps evolutionary—link between kin recognition, historecognition and the enhancement of colony fusion has been experimentally demonstrated. The prevalence of limited dispersal⁹ and historecognition systems among sessile clonal marine invertebrates³² suggests that such a link may be found in other clonal organisms. Recent studies of congeneric mice show that haplotypic differences in the H-2 region affect mating behaviour⁴⁵; thus, histocompatibility markers could mediate kin recognition in other taxa. Whether the study of marine invertebrates will provide useful insights into the more complex dynamics of kin recognition in mobile, social vertebrates is unclear, but the analogies remain intriguing.

This research was funded by NSF grant OCE 84-07158 to R.K.G. The Committee on Research of the University of California at Davis provided additional support. We thank P. Sherman and M. Marthas for their useful comments.

Received 27 January; accepted 25 June 1986.

- Hamilton, W. D. *J. theor. Biol.* **7**, 1–51 (1964).
- Holmes, W. G. & Sherman, P. W. *Am. Zool.* **22**, 491–517 (1982).
- Shields, W. J. *Philopatry, Inbreeding and the Evolution of Sex* (State University of New York Press, Albany, 1982).
- Sherman, P. W. & Holmes, W. G. *Fortschr. Zool.* **31**, 437–459 (1985).
- Keough, M. J. *Evolution* **38**, 142–147 (1984).
- van Duyl, F. C., Bak, R. P. M. & Sybesma, J. *Mar. Ecol. Prog. Ser.* **6**, 35–42 (1981).
- Olson, R. R. *Ecology* **66**, 30–39 (1985).
- Jackson, J. B. C. in *Population Biology and Evolution of Clonal Organisms* (eds Jackson, J. B. C., Buss, L. W. & Cook, R. E.) 297–356 (Yale University Press, New Haven, 1986).
- Jackson, J. B. C. *J. mar. Res.* (in the press).
- Kott, P. *Proc. 2nd int. Coral Reef Symp.* **1**, 405–423 (1974).
- Grave, C. & Woodbridge, H. J. *Morph. Physiol.* **39**, 207–247 (1924).
- Clark, P. J. & Evans, F. C. *Ecology* **35**, 445–453 (1954).
- Simberloff, D. S. *Ecology* **60**, 679–685 (1979).
- Efron, B. & Gong, G. *Am. Statist.* **37**, 36–48 (1983).
- Bancroft, F. W. *Proc. Calif. Acad. Sci.* **3**, 138–186 (1903).
- Oka, H. & Watanabe, H. *Bull. biol. Sin. Asamushi* **10**, 153–155 (1960).
- Sabbadin, A. *Atti. Accad. naz. Lincei R.* **32**, 1031–1035 (1962).

18. Mukai, H. & Watanabe, H. *Proc. Japan Acad.* **51**, 44-47 (1975).
19. Karakashian, S. & Milkman, R. *Biol. Bull.* **133**, 473 (1967).
20. Curtis, A. S. G., Kerr, J. & Knowlton, N. *Transplantation* **33**, 116-123 (1982).
21. Bak, R. P. M., Sybesma, J. & van Duyl, F. C. *Mar. Ecol. Prog. Ser.* **6**, 43-52 (1981).
22. Russ, G. R. *Oecologia* **53**, 12-19 (1982).
23. Reiswig, H. R. *Bull. mar. Sci.* **23**, 191-226 (1973).
24. Fry, W. G. in *4th Eur. mar. biol. Symp.* (ed. Crisp, D. J.) 155-178 (Cambridge University Press, 1971).
25. Highsmith, R. C., Riggs, A. C. & D'Antonio, C. A. *Oecologia* **46**, 322-329 (1980).
26. Wahle, C. M. *Biol. Bull.* **165**, 778-790 (1983).
27. Hughes, T. P. & Jackson, J. B. C. *Ecol. Monogr.* **55**, 141-166 (1985).
28. Sebens, K. P. *J. exp. mar. Biol. Ecol.* **72**, 263-285 (1983).
29. Buss, L. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5355-5359 (1980).
30. Winston, J. E. & Jackson, J. B. C. *J. exp. mar. Biol. Ecol.* **76**, 1-21 (1984).
31. Fadlallah, Y. H. *Coral Reefs* **2**, 129-150 (1983).
32. Buss, L. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5337-5341 (1982).
33. Jones, W. E. *Nature* **178**, 426-427 (1956).
34. Kaufmann, K. W. *Oecologia* **49**, 293-299 (1981).
35. Hidaka, M. *Coral Reefs* **4**, 111-116 (1985).
36. Buss, L. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1387-1391 (1983).
37. Nieuwkoop, P. D. & Sutasurya, L. A. *Primordial Germ Cells in the Invertebrates* (Cambridge University Press, 1981).
38. Sabbadin, A. & Zaniolo, G. *J. exp. Zool.* **207**, 289-304 (1979).
39. Buss, L. W. & Green, D. R. *Dev. comp. Immun.* **9**, 191-201 (1985).
40. Sabbadin, A. *Dev. Biol.* **24**, 379-391 (1971).
41. Oka, H. in *Profiles of Japanese Science and Scientists* (ed. Yukawa, H.) 195-206 (Kodansha, Tokyo, 1970).
42. Scofield, V. L., Schlumpberger, J. L., West, L. A. & Weismann, I. L. *Nature* **295**, 499-502 (1982).
43. Grosberg, R. K. *Evolution* (submitted).
44. Strathmann, R. R., Strathmann, M. & Emson, R. H. *Am. Nat.* **123**, 796-818 (1984).
45. Yamazaki, K. E., Beauchamp, G. K., Bard, J., Thomas, L. & Boyse, E. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7828-7831 (1982).

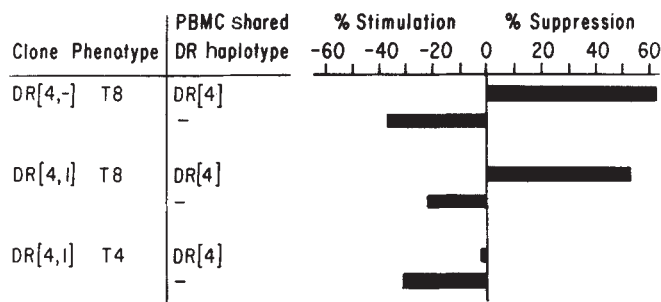


Fig. 1 HLA-D-restricted lepromin-induced suppression of Con A responses of normal peripheral blood mononuclear cells (PBMC) by T8 clones from lepromatous leprosy skin lesions.

Methods. Ellipsoid skin biopsy specimens obtained under local anaesthesia (1% lidocaine with adrenaline) from lepromatous leprosy patients, classified according to the criteria of Ridley and Jopling³⁰, were stripped of epidermis and subcutaneous fat and placed in a tissue sieve fitted with a 64-µm-mesh filter (Bellco Glass Inc.) on a Petri dish containing RPMI 1640 medium with 10% fetal calf serum¹². The tissue was cut into small pieces with a surgical scalpel and extruded through the mesh with a glass rod. The lymphocytes were isolated by Ficoll-Paque (Pharmacia) centrifugation, labelled with fluorescein isothiocyanate-conjugated monoclonal antibodies T4 and T8 (Coulter Immunology), and sorted using a FACS IV (Becton-Dickinson). Long-term culture lines were successfully established from T lymphocytes derived from skin lesions of lepromatous patients. T4 and T8 lymphocytes sorted by FACS to greater than 90% purity were seeded in 96-well flat-bottom Linbro microtitre plates (Flow Laboratories) at 1,000 cells per well. To each well were added 5×10^5 lethally irradiated (5,000 rad) allogeneic feeder peripheral blood mononuclear cells, RPMI 1640 with penicillin/streptomycin, 10% human AB serum and 10% IL-2 (Electronucleonics). Cultures were incubated at 37°C in 5% CO₂. Every 3 or 4 days thereafter, the cultures were monitored with an inverted phase microscope, transferred to additional wells when necessary and fed with 50% fresh IL-2-containing medium. To avoid suppression from adherent suppressor cells found in blood of lepromatous patients, the lines were cultured with allogeneic or no feeder cells. Some lines were maintained for longer than 11 months using these techniques. Lines from two lepromatous patients (J.G. and B.P.) were selected for subcloning. Clones were obtained by limiting dilution by seeding 0.3 cells per 20 µl Terasaki well with 10% AB serum, 5-10% IL-2 and 10^4 γ-irradiated allogeneic feeder PBMC for 1 week¹³. Positive wells were transferred to a 96-well plate and expanded as above, using 5×10^4 feeder cells, and then to 24-well plates using 5×10^5 feeder cells. When ready for study, T-lymphocyte clones were separated from non-viable cells over Ficoll-Paque and the phenotype was determined by FACS. Antibodies used included T4, T8, anti-HLA-DR (Becton-Dickinson) and anti-IL-2 receptor (Tac, kindly provided by Dr T. Waldmann, or from Becton-Dickinson). Measurement of antigen-induced suppressor activity of lesion derived T-lymphocyte clones was performed using the method of Mehra *et al.*³. 5×10^4 Cells of a particular T8 clone were admixed with 2×10^5 normal PBMC per microtitre well with and without Dharmendra lepromin (1:10). Con A ($2.5 \mu\text{g ml}^{-1}$) was added and ³H-thymidine incorporation of triplicate cultures measured at 3 days. Per cent suppression is calculated as $100 - (\text{c.p.m. Con A} + \text{lepromin} / \text{c.p.m. Con A}) \times 100$. For each clone, three HLA-DR-matched (A.K., M.R., H.K.) and three HLA-DR-mismatched (B.L., G.P., M.S.) donors were used and the results were averaged. A non-reactive T4 clone derived from a lepromatous lesion was substituted for the T8 clones to control for the addition of cells. The c.p.m. above background obtained in the Con A controls averaged 27,471. The results obtained with the T8 clones admixed with the HLA-DR-matched cells differed significantly ($P < 0.001$) when compared with either the T8 clones admixed with HLA-DR-mismatched cells or the T4 'filler' clone assayed with HLA-DR-matched cells, using the z-test approximation to the test for relative difference in Poisson data on actual c.p.m.

(Leu 2a) phenotype were present in excess in lepromatous granulomata as compared with tuberculoid⁷⁻¹⁰. Because the surface markers do not necessarily reflect lymphocyte function within lesions¹¹, simple procedures were developed to extract 10^5 - 10^6 lymphocytes from biopsies of leprosy skin lesions¹². The phenotypic distribution of these cells as determined by fluorescence-activated cell sorter (FACS) analysis was similar to that obtained using immunoperoxidase staining of the tissue sections (T4:T8 = 0.5:1 in lepromatous and T4:T8 = 2.0:1 from tuberculoid lesions). In lepromatous patients, the T4/T8 ratio of extracted cells was significantly different from that in blood, indicating that contamination from peripheral blood was negligible. Because we assumed that the lymphocytes in the lesions may be involved in specific immune reactivity and were likely

Genetically restricted suppressor T-cell clones derived from lepromatous leprosy lesions

Robert L. Modlin[†], Hideyuki Kato[‡], Vijay Mehra[‡], Erica E. Nelson[†], Fan Xue-dong[‡], Thomas H. Rea^{*}, Paul K. Pattengale[†] & Barry R. Bloom[‡]

Section of * Dermatology and † Department of Pathology, University of Southern California School of Medicine, Los Angeles, California 90033, USA

‡ Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York 10461, USA

Leprosy is a spectral disease in which immune responses to *Mycobacterium leprae* correlate with the clinical, bacteriological and histopathological manifestations of disease^{1,2}, so study of its pathology provides insights into immunoregulatory mechanisms in man. At the tuberculoid pole, patients have few lesions in the skin which contain rare organisms and are able to mount strong cell-mediated immune responses to *M. leprae* antigens. In contrast, at the lepromatous pole, patients have disseminated skin lesions containing large numbers of acid-fast bacilli and are selectively unresponsive to antigens of *M. leprae*. *M. leprae*-induced suppressor cells derived from peripheral blood have been reported to be active *in vitro*³⁻⁶, yet their *in vivo* significance has remained unclear. Because the focal point of the immune response to *M. leprae* is the skin lesion consisting of lymphocytes and macrophages, we have recently developed methods for isolating lymphocytes from skin biopsies of leprosy patients. We report here that two T8 clones derived from lepromatous leprosy skin biopsies, in the presence of lepromin, suppress concanavalin A (Con-A) responses both of peripheral blood mononuclear cells and of T4 clones in an HLA-D (HLA, histocompatibility locus antigen)-restricted manner. Moreover, these T8 clones suppressed responses of HLA-D-matched, but not HLA-D-mismatched antigen-responsive T4 clones to *M. leprae* antigens, indicating that T-cell suppression is major histocompatibility complex (MHC)-restricted at some level in man.

Initial immunohistological studies of skin lesions from patients across the spectrum revealed that cells bearing the T8

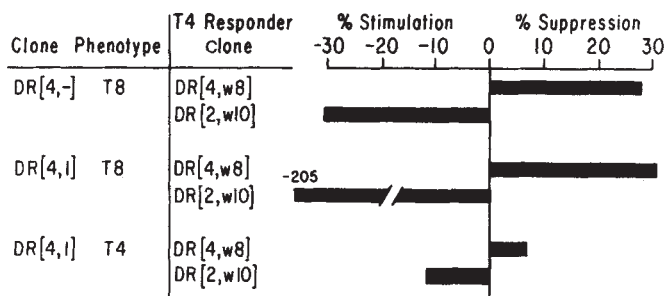


Fig. 2 HLA-D-restricted lepromin-induced suppression of Con A responses of T4 clones by T8 clones from lepromatous leprosy skin lesions.

Methods. Two antigen-reactive T4 clones were derived from skin and blood respectively of two tuberculoid leprosy patients, O.C. and I.B. A T4 line was obtained from peripheral blood by stimulating 1×10^6 PBMC ml^{-1} with Dharmendra lepromin (1:10) and 10% AB serum. On day 5, 10% IL-2 was added. On day 6 lymphoblasts were isolated on a Ficoll gradient, then recultured with lepromin, IL-2 and feeder cells. A T4 line was obtained from a tuberculoid skin lesion by extracting T4 cells as described in Fig. 1 legend, and culturing in the presence of lepromin and IL-2. The T4 line from skin was cloned by picking colonies at 4 days from a 35-mm Petri dish containing 1% methylcellulose in RPMI with lepromin (1:10), 10% AB serum, 10% IL-2, 10^2 – 10^4 lymphoblasts and 4×10^6 irradiated autologous feeder cells per ml. These two T4 lines had previously been found to respond to lepromin but not tetanus toxoid. The T4 line from blood was cloned by limiting dilution as described except that lepromin was included in the cultures at a 1:10 dilution. For testing, 10^4 cells of an antigen-reactive T4 clone were mixed with 10^5 irradiated autologous PBMC. Lepromin-induced suppression of the Con A response of these T4 clones was determined at 72 h in the presence or absence of 10^4 cells of the T8 clones. A non-reactive T4 clone was substituted for the T8 clone to control for the addition of cells. The mean c.p.m. above background in three experiments for T4 clone O.C. was 19,059 c.p.m. and for the I.B. clone was 18,150 following Con A stimulation. The results obtained with the T8 clones when admixed with the HLA-DR-matched antigen-reactive T4 clone were significantly different ($P < 0.001$ by z -test) compared with either the T8 clones when assayed with the mismatched antigen-reactive T4 clone or the control T4 'filler' clone admixed with the HLA-DR-matched antigen-reactive T4 clone.

to be activated *in situ*, cells from lesions were isolated, sorted for T4 and T8 cells and expanded in culture as long-term (up to 11 months) lines in the presence of irradiated feeder cells and interleukin-2 (IL-2) (ref. 12). The cells were cloned by limiting dilution (0.3 cells per well) as described by Lamb *et al.*¹³. T4 clones were grown in the presence of lepromin (a human-derived *M. leprae* preparation) and IL-2 plus autologous or HLA-D-matched feeder cells. T8 clones from lepromatous lesions were grown in the presence of γ -irradiated feeder cells and IL-2 but not antigen. Lepromin seemed to be inhibitory to the growth of T8 clones. The T-cell lines and clones studied expressed both HLA-DR monomorphic determinants and the Tac receptor for IL-2. The T8 clones grow rather slowly and have been difficult to expand in culture. Nevertheless, two clones were selected for study which were suppressive only on stimulation with antigen as described; they have been in culture for 12 and 6 months, respectively.

Because it has been previously observed that lepromatous patients have circulating antigen-specific suppressor T8 cells that inhibit mitogen-induced T-cell proliferation *in vitro*^{3,4}, lines containing predominantly T8 cells were initially screened for lepromin-induced suppression of the Con A responses of the peripheral blood mononuclear cells (PBMC) of normal donors¹². Lepromin-specific suppression of PBMC from random normal donors was observed in 13 of 32 T8 lines derived from

skin biopsies of 6 lepromatous leprosy patients and in none of 9 T8 lines derived from 6 tuberculoid lesions. The inconsistent suppression of T8 lines from the lesions prompted us to select for single-cell clones and consider possible HLA restriction of suppression.

The HLA types of the donors whose cells were studied are listed in Table 1. The ability of cloned T8 cells from two HLA-DR4 donors to suppress the Con A responses of a panel of six DR4-matched and -mismatched normal donors' PBMC was examined. As illustrated in Fig. 1, there was marked lepromin-induced suppression by the T8 clones of Con A responses of primary lymphocytes from donors sharing DR4 and not those mismatched for DR. There was no obvious correlation with HLA-A or -B-types in these combinations. A Tac⁺, IL-2-dependent T4 clone derived from lepromatous skin of one of the same DR4 donors was used in every experiment to control both for the effect of addition of cloned cells to the wells and for possible competition for IL-2 in the medium. In no case was suppression observed in this control.

Two lepromin-specific T4 clones were derived from blood and skin respectively of two strongly lepromin skin-test-positive patients of differing DR types with tuberculoid leprosy. The ability of the T8 clones to suppress the responses of these T4 clones to mitogen and specific antigen was tested. When DR4-matched and DR-mismatched T4 clones and T8 clones were compared for lepromin-induced suppression of Con A responses at 72 hours, again there was striking suppression only of HLA-D-compatible cultures (Fig. 2). Finally, since the relevance of the lepromin-induced suppression of mitogenic activity to the failure of lepromatous patients to respond to specific antigens remained to be established, we examined the effect of the T8 clones on proliferation of the T4 clones to specific antigen. Since the T4 clones proliferate in response to lepromin at 72 h and fail to show any alloreactivity to MHC-incompatible cells at that time, we were in a unique position to examine whether the T8 clones were capable of suppressing the proliferative response of T4 clones to specific antigen, and to ascertain whether such suppression was MHC-restricted. The results of three experiments (Fig. 3) revealed an HLA-D-restricted suppression by T8 clones of the responses of T4 clones to *M. leprae* antigens.

While some previous efforts to detect antigen-specific suppression by lymphocytes from lepromatous leprosy patients have not been successful^{14,15}, Ottenhof *et al.*¹⁶ have recently obtained data similar to those presented here indicating that T8 clones derived from blood of lepromatous patients are able to suppress specific responses of T4 clones from the same patients to lepromin. Nishimura and Sasazuki, who reported the suppression of streptococcal antigen responses of HLA-D-matched T cells by T8 cells from hyporesponsive donors¹⁷, have recently observed suppression of lepromin responses by T8 cells from lepromatous patients (Kikuchi, I., Uzawi, T. and Sasazuki, T.,

Table 1 HLA phenotypes of donor cells

	Donor	HLA-A	HLA-B	HLA-DR
T8 clones	J.G.	24, 32	12, 18	4, —
	B.P.	11, 29	7, 35	4, 1
T4 clones	I.B.	24, 31	61, 62	4, w8
	O.C.	11, 24	38, —	2, w10
Peripheral blood mononuclear cells	A.K.	2, 11	44, 60	4, —
	M.R.	28, —	38, 60	4, —
	H.K.	24, —	61, —	4, w9
	B.L.	26, 28	35, 51	w6, —
	G.P.	1, 2	8, 60	2, —
	M.S.	11, 23	45, 44	2, —

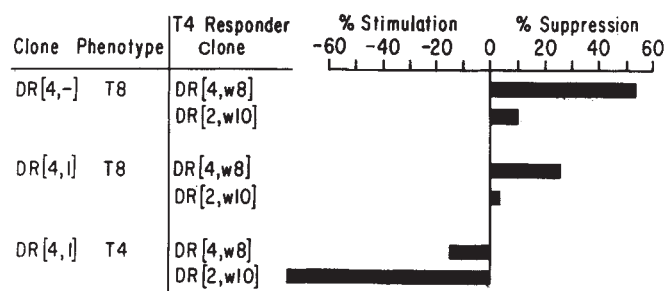


Fig. 3 HLA-D-restricted suppression of the antigen response of T4 clones by T8 clones. Antigen-reactive T4 cloned cells (10^4) were cultured with 10^5 irradiated autologous PBMC per 200- μ l micro-well in RPMI with 10% AB serum, with and without lepromin (1:10), and with and without 10^4 of each T8 clone. Thymidine incorporation was measured at 72 h. Per cent suppression was determined as $100 - (\text{c.p.m. T4 clone} + \text{T8 clone} / \text{c.p.m. T4 clone}) \times 100$. A non-reactive T4 clone was substituted for the T8 clone to control for the addition of cells. The mean c.p.m. above background for the lepromin-stimulated T4 clones was 33,662 c.p.m. for I.B. and 24,452 c.p.m. for O.C. respectively. The results obtained with the T8 clones when admixed with the HLA-DR-matched antigen-reactive T4 clone were significantly different ($P < 0.001$ by z-test) compared with either the T8 clones assayed with the mismatched antigen-reactive T4 clone or the control T4 'filler' admixed with the antigen-reactive T4 clone.

in preparation). There are previous reports of T8 suppressor cell clones from human peripheral blood^{18,19}, but we believe that the present observations represent the first evidence that, at some level, human T-cell suppression may be HLA-D-restricted. More detailed analysis will be required to establish that the restriction is in fact localized to the HLA-D region and to learn whether DP, DQ or DR are involved. There is evidence from murine systems that effector suppression is mediated by Ia-positive T cells which are MHC class II-restricted²⁰⁻²².

The nature of the antigen recognized by the suppressor T8 clones remains to be established. We have previously reported that some T-suppressor cells appear to recognize the unique phenolic glycolipid of *M. leprae*²³. The recent successful expression of *M. leprae* antigens in *Escherichia coli*²⁴ should permit the delineation of protein antigens and epitopes recognized by T-suppressor and T-helper cells.

The existence of monocytes in lepromatous patients with the ability to suppress mitogen and antigen responses *in vitro* has been reported by a number of laboratories^{3,5,6,25,26}. The non-specific nature of the monocyte suppression and the selective unresponsiveness of lepromatous patients to antigens of *M. leprae* suggest that more specific mechanisms must be involved as well. A number of lines of evidence suggest that T-suppressor cells may contribute to the failure of the immune system in lepromatous patients to kill and clear *M. leprae* from their lesions. Clearly, large numbers of *M. leprae* coexist in lesions with T8 cells which show antigen-induced suppressor activity *in vitro*. There is a marked diminution of T4 cells containing IL-2 in these same lesions, although many are positive for the IL-2 (Tac) receptor²⁷. We have found that the *in vitro* suppression of Con A responses by T8 cells from lepromatous patients is specific for the stage of the disease^{3,4}. Finally, T8 suppressor activity present in 10 lepromatous patients disap-

peared after successful immunotherapy by vaccination with live BCG together with killed *M. leprae*²⁸.

At least four possible mechanisms could be involved in the selective immunological unresponsiveness of lepromatous leprosy patients: (1) *M. leprae* or a secreted metabolic product could inhibit appropriate antigen presentation to T4 cells by macrophages, Langerhans or dendritic cells, resulting in development of T8 suppressor cells; (2) T8 suppressor cells could secrete a suppressor factor (TsF) that blocks responsiveness of specific T4 cells to *M. leprae* antigens; (3) T8 suppressor cells could lyse T4 cells in an antigen- and MHC-restricted manner; and (4) T8 cells could interact with and 'veto' antigen-reactive T4 cells without killing them²⁹.

We believe this to be the first report of isolation of genetically restricted T-suppressor cell clones from lesions of a human disease. Considerably more detailed study will be required to establish the precise nature of the genetic restrictions involved in unresponsiveness in man. Nevertheless, we hope that this approach will be useful in exploring the nature of genetic restrictions and receptor gene organization in suppressor T cells, and the role of these cells in regulating immunological unresponsiveness in man.

This work was supported by NIH grants AI 22187, AI 22553, AI 07118 and AI 02111, the UNDP/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases (IMMLEP), National Hansen's Disease Center, Knights of St Lazarus of Jerusalem, Heiser Trust, Hartford Foundation, the US-Japan Leprosy Program, and the Irvington House Institute for Medical Research. P.K.P. is a Scholar of the Leukaemia Society. We thank Drs Jonathan Lamb, Geoffrey Gersuk and Roland Tchong for help and advice on T-cell cloning, Ms Betty Donovan for secretarial assistance and Bryan Langholz for help with statistical analysis.

Received 16 December 1985; accepted 22 April 1986.

- Rea, T. H. & Levan, N. E. *Archs Derm.* **113**, 345-352 (1977).
- Bloom, B. R. & Godal, T. *Rev. infect. Dis.* **5**, 765-779 (1983).
- Mehra, V., Mason, L. H., Fields, J. P. & Bloom, B. R. *J. Immun.* **123**, 1813-1817 (1979).
- Bloom, B. R. & Mehra, V. *Immun. Rev.* **80**, 6-28 (1984).
- Salgame, P. R., Birdi, T. J., Mahadevan, P. R. & Antia, N. H. *Int. J. Lepr.* **48**, 172-177 (1980).
- Nath, I., Van Rood, J. J., Mehra, N. K. & Vaidya, M. C. *Clin. exp. Immun.* **42**, 203-210 (1980).
- Van Voorhis, W. C. *et al. New Engl. J. Med.* **307**, 1593-1597 (1982).
- Modlin, R. L., Hofman, F. M., Taylor, C. R. & Rea, T. H. *J. Am. Acad. Derm.* **8**, 181-189 (1983).
- Modlin, R. L. *et al. Clin. exp. Immun.* **51**, 430-438 (1983).
- Narayanan, R. B., Bhutani, L. K., Sharma, A. K. & Nath, I. *Clin. exp. Immun.* **51**, 422-429 (1983).
- Modlin, R. L. *et al. Clin. exp. Immun.* **60**, 241-248 (1985).
- Modlin, R. L. *et al. J. Immun.* (in the press).
- Lamb, J. R. *et al. J. Immun.* **128**, 233-238 (1982).
- Stoner, G. L., Touw, J., Belehu, A. & Naafs, B. *Lancet* **ii**, 543-547 (1978).

- Nath, I., Van Rood, J. J., Mehra, N. K. & Vaidya, M. C. *Clin. exp. Immun.* **42**, 203-210 (1980).
- Ottenhoff, R. H. M., Elferink, D. G., Klatser, P. R. & de Vries, R. R. P. *Nature* (in the press).
- Nishimura, Y. & Sasazuki, T. *Nature* **302**, 67-69 (1983).
- Bensussan, A., Acuto, O., Hussey, R. E., Milanese, C. & Reinherz, E. L. *Nature* **311**, 565-567 (1984).
- Rees, A. D. M. *et al. in Human T Cell Clones, a New Approach to Immunoregulation* (eds Feldmann, M., Lamb, J. R. & Woody, J. N.) (Humana, Clifton, New Jersey, 1985).
- Baxevanis, C. N., Ishii, N., Nagy, Z. A. & Klein, J. J. *Exp. Med.* **156**, 822-833 (1982).
- Dorf, M. E. & Benacerraf, B. *Immun. Rev.* **83**, 23-40 (1985).
- Araneo, B. A. & Yowell, R. L. *J. Immun.* **135**, 73-79 (1985).
- Mehra, V., Brennan, P. J., Rada, E., Convit, J. & Bloom, B. R. *Nature* **308**, 194-196 (1984).
- Young, R. A. *et al. Nature* **316**, 450-452 (1985).
- Salgame, P. R., Mahadevan, P. R. & Antia, N. H. *Infect. Immunity* **40**, 1119-1126 (1983).
- Sathish, M., Bhutani, L. K., Sharma, A. K. & Nath, I. *Infect. Immunity* **42**, 890-899 (1983).
- Modlin, R. L. *et al. J. Immun.* **132**, 3085-3090 (1984).
- Mehra, V., Convit, J., Rubinstein, A. & Bloom, B. R. *J. Immun.* **129**, 1946-1951 (1982).
- Miller, R. G. *Immun. Today* **7**, 112-114 (1986).
- Ridley, D. S. & Jopling, W. H. *Int. J. Lepr.* **34**, 255-273 (1966).

Cloned suppressor T cells from a lepromatous leprosy patient suppress *Mycobacterium leprae* reactive helper T cells

Tom H. M. Ottenhoff*, Diënné G. Elferink*, Paul R. Klatser† & René R. P. de Vries*

* Department of Immunohaematology and Bloodbank, University Hospital Leiden, 2333 AA Leiden, The Netherlands

† Laboratory of Tropical Hygiene, Royal Tropical Institute, 1105 AZ Amsterdam, The Netherlands

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*. A characteristic feature of the disease is its remarkable spectrum of clinical symptoms correlating with the cellular immune responsiveness of the patient¹. At one pole of this spectrum are tuberculoid patients displaying both acquired cell-mediated immunity and delayed type hypersensitivity against the bacillus²⁻⁴. At the other pole are lepromatous patients which show a specific T-cell unresponsiveness against *M. leprae*⁵. In between those two poles variable degrees of tuberculoid and lepromatous features may be seen in borderline leprosy patients. Thus far, studies on the mechanism of the antigen specific unresponsiveness in lepromatous leprosy have been contradictory and difficult to interpret, probably because of the use of heterogeneous cell populations in those experiments⁶⁻¹⁰. We have now succeeded in cloning *M. leprae* stimulated T-helper (T_H) as well as T-suppressor (T_S) cells from a borderline lepromatous patient. The T_S-clones of this patient specifically suppress responses of peripheral T_H cells as well as T_H clones induced by both *M. leprae* and other mycobacteria, but not unrelated antigen or mitogen. These T_S cells also completely suppress T_H cell responses against a *M. leprae* specific protein with a relative molecular mass of 36,000 (36K), suggesting the presence of a suppression inducing determinant on this 36K *M. leprae* protein.

In contrast to *M. leprae*-activated and interleukin-2 (IL-2)-propagated T-cell lines and clones derived from tuberculoid leprosy patients, similar T-cell lines of (borderline) lepromatous leprosy patients consistently failed to show a proliferative response against *M. leprae* antigens presented by autologous or allogeneic histocompatibility locus antigens (HLA) class II matched antigen presenting cells (APC) (data not shown). This lack of proliferation by cultured T cells to *M. leprae* was also observed when the peripheral blood mononuclear cells (PBMNC) of such (borderline) lepromatous patients did proliferate to *M. leprae*. We selected one such a borderline lepromatous patient whose peripheral T cells did proliferate against *M. leprae* as well as to unrelated antigens like herpes simplex virus (HSV) and mitogen (phytohaemagglutinin; PHA) because this might enable us to study both helper and possibly suppressor T-cell responses against *M. leprae*. We then tested the *M. leprae*-non-responsive T-cell line on autologous antigen activated peripheral T_H cells, thereby circumventing allogeneic effects. As shown in Fig. 1, the *M. leprae*-non-responsive T-cell line suppressed specifically and strongly *M. leprae* but not HSV or PHA responses. Peripheral responding T cells were mainly positive for the CD4 antigen. The same result was obtained with five other *M. leprae* induced T cell lines of the same patient, all of which were antigen non responsive but IL-2 responsive.

Cloned T cells were derived from one such a line. None of these clones proliferated to *M. leprae* in the presence of APC, similar to the parental T-cell line (data not shown). All these clones suppressed *M. leprae* but not HSV-specific responses of peripheral T cells (Table 1). Peripheral T-cell responses against *M. tuberculosis*, *M. fortuitum* or PPD (purified protein derivative of *M. tuberculosis*) were also suppressed in variable degrees by

these T_S-clones. The cell surface marker phenotypes of the T_S-cell line and T_S clones are also shown in Table 1. All T_S-clones were CD3 positive, confirming the T-cell lineage nature of these cells. All T_S clones were also CD8 positive, whereas some showed a weak (1D8, 1D11) or clear (1E9) double staining for CD4. The simultaneous expression of CD4 and CD8 by peripheral T cells and T-cell clones has been reported recently^{11,12}. The clones strongly expressed HLA-DR antigens whereas the expression of DQ and DP varied.

Recently the genes for the five major proteins of *M. leprae* have been cloned and expressed in *Escherichia coli*¹³. One of these five proteins is a *M. leprae* specific 36K protein¹³⁻¹⁵. This protein was recently shown to contain several different antigenic determinants capable of stimulating helper T cells. These determinants include both *M. leprae* specific and cross-reactive ones as defined by *M. leprae* specific and cross reactive T-helper clones from tuberculoid patients¹⁶. Peripheral T cells of the suppressor cell donor reacted with this 36K protein. This offered the possibility to determine whether these 36K reactive T_H cells could be suppressed by the autologous T_S cells. A negative result would indicate that the epitope(s) recognized by T_S cells resides neither on the 36K protein nor on the 36K reactive T_H cells, thus enabling us to dissociate easily the induction of *M. leprae* reactive T_H and T_S. This is of course not only important for experimental purposes, but also for the prevention (vaccine) and maybe even therapy of leprosy. However, as shown in Fig. 2, the 36K response could be inhibited completely by the T_S cells. This result may indicate that the 36K protein in addition to helper epitopes also contains at least one suppressor epitope.

Table 1 T_S clones and line suppress the response of PBMNC to *M. leprae* and other mycobacteria but not unrelated antigen

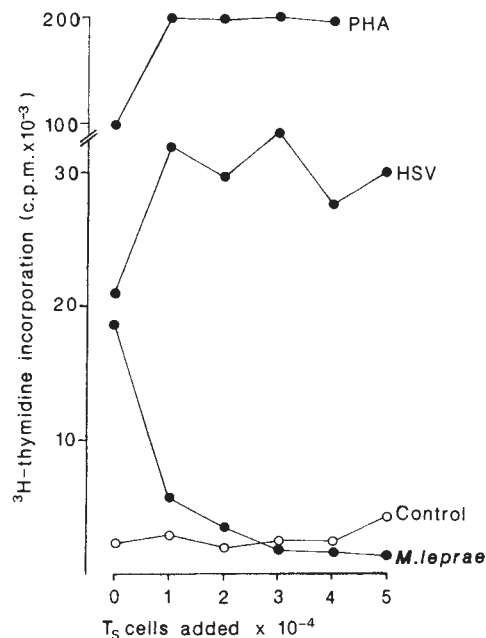
Ts clone [CD phenotype]	1D8 [4,8]	1D11 [4,8]	1E9 [4(8)]	1D10 [8]	3D2 [8]	1G2 [8]	Ts line [4,8]
% Suppression of response of PBMNC to:							
<i>M. leprae</i>	70	95	73	53	54	53	88
<i>M. tuberculosis</i>	24	55	70	44	39	ND	80
<i>M. fortuitum</i>	20	48	72	34	40	ND	91
PPD	2	44	56	50	50	ND	76
HSV	-250	12	-8	-231	-56	5	-59

The results are expressed as the percentage suppression of peripheral T cell responses against the respective antigens after the addition of 3×10^4 T suppressor cells, as calculated with the formula $(1 - \text{c.p.m. of peripheral T cells cultured in the presence of suppressor T cells} / \text{c.p.m. of peripheral T cells in the absence of suppressor T cells}) \times 100\%$. Peripheral T-cell responses in the absence of suppressor cells against *M. leprae* were: 21,920 c.p.m., against *M. tuberculosis*: 26,570, against *M. fortuitum*: 15,605, against PPD: 6,060 and against HSV: 29,300. Responses in the absence of antigen were 970 c.p.m. All results are given as the mean c.p.m. of duplicate cultures. Standard deviations did not exceed 20%. The CD4,8 phenotypes of the T cell line and clones are shown. ND, not determined.

Methods. The cultures were set up as described in Fig. 1 legend. The antigens tested in addition were ultrasonicates of whole mycobacteria ($20 \mu\text{g ml}^{-1}$) and PPD ($5.0 \mu\text{g ml}^{-1}$). From the suppressor T-cell line, clones were prepared by cloning the cells (1D8, 1D10, 1D11: 10 cells per well; 1E9, 3D2, 1G2: 0.5 cell per well) on the feeder cell mixture described in the legend of Fig. 1, supplemented with Leuko Agglutinin A ($1 \mu\text{g ml}^{-1}$) (Pharmacia) and expanding the cultures in the presence of exogenous IL-2. This cycle was repeated several times¹⁷. The cell surface marker phenotypes of the different T cells were determined by a standard indirect immunofluorescence technique as described in ref. 32. The monoclonal antibodies used were OKT3 (Ortho) for the CD3 marker, RIV-6 (anti-T4/leu 3; National Institute of Public Health, Bilthoven, Netherlands) for the CD4 marker, FK18 (anti-T8/leu2; gift of F. Koning) for the CD8 marker, B8.11.2 (anti HLA-DR; gift of B. Malissen), SPV-L3 (anti HLA-DQ; gift of H. Spits) and B7/21 (anti HLA-DP; gift of F. Bach) for respectively HLA-DR, DQ and DP antigens.

Fig. 1 Antigen specific suppression of *M. leprae* but not herpes simplex virus (HSV) or PHA induced proliferative peripheral T cell responses by a *M. leprae* induced and in IL-2 grown suppressor T cell line. The latter was added to 10^5 PBMNC in concentrations of 0, 1, 2, 3, 4 or 5×10^4 cells per culture. The results are expressed as the mean c.p.m. of duplicate cultures (^3H -thymidine incorporation). Standard deviations mostly did not exceed 20%. *M. leprae* antigens isolated from human lepromas and from armadillo infected tissue induced similar suppression.

Methods. PBMNC of a borderline lepromatous leprosy patient were isolated from heparinized venous blood by Ficoll-Isopaque density centrifugation (Pharmacy, University Hospital Leiden), washed three times in Hanks' balanced salt solution (Gibco) and resuspended in Iscove's modified Dulbecco's medium (IMDM; Gibco) supplemented with $100 \mu\text{g ml}^{-1}$ streptomycin, 100 U ml^{-1} penicillin (both Flow Laboratories) and 10% heat inactivated human AB serum. The diagnosis of this and other patients used for this study was based on regular clinical examination, lepromin skin test and skin biopsy histology by Dr D. L. Leiker, Department of Dermatology, University Hospital of Amsterdam, the Netherlands. A standard lymphocyte transformation test was set up with 10^5 PBMNC per culture, to which antigen was added. Antigens tested were Dharmendra lepromin, consisting of bacilli isolated from human lepromas (obtained from the Centre for Infectious Diseases, Center for Disease Control, Atlanta, USA; 1:80 dilution) and herpes simplex virus (obtained from the National Institute of Public Health, Bilthoven, The Netherlands; 1:64 dilution). From the start of the culture, either no or 1, 2, 3, 4 or 5×10^4 T cells of the nonresponsive T-cell line (see above) were added to the culture. The cultures were set up in flat-bottomed 96-well microtitre plates (Greiner), incubated at 37°C in a fully humidified 5% CO_2 -air mixture for 5 days after which $1.0 \mu\text{Ci } ^3\text{H}$ -thymidine was added to each well; the cultures were collected 16 h later on glass-fibre filters using a semiautomatic sample harvester. ^3H -TdR incorporation was assessed by liquid scintillation counting. The T cell line was generated as described previously^{16,30}. In brief, PBMNC were restimulated in vitro with an optimal concentration of Dharmendra lepromin (1:80) in 24-well tissue culture plates. T-cell blasts were then cultured for 10 days in the presence of 20% IL-2 containing medium (Lymphocult-T, Biotest). The cells were tested then for antigen-induced proliferation in the presence of autologous APC and antigen but failed to show a response. The cells were then restimulated with a feeder mixture consisting of irradiated autologous EBV-BC as autologous APC³¹ (10^5 cells ml^{-1}), PBMNC from three or four random donors (10^6 cells ml^{-1}) and Dharmendra lepromin (1:80). The cultures were further expanded in the presence of IL-2. The cells were then collected, tested again for non responsiveness against antigen and used for experiments. All cells were cultured in IMDM supplemented with 10% human AB serum.



This 36K epitope then would be the first *M. leprae* suppressor epitope on a *M. leprae* protein, because we consider it unlikely that the 36K protein would contain the *M. leprae* specific phenolic glycolipid, which has been reported to induce suppression of concanavalin-A stimulated PBMNC of lepromatous patients^{5,7}. The co-existence of both helper and suppressor epitopes on the 36K protein would closely resemble the situation for other immunogenic proteins like hen egg white lysozyme¹⁷, β -galactosidase¹⁸ and a 185K streptococcal protein¹⁹. For the first two protein antigens it has been shown that the putative suppressor epitope has to be situated on the same molecule or fragment as the helper epitope in order to induce suppression of the immune response^{20,21}. This may be true for the 36K protein as well. A second explanation however for the suppression of T_H cell responses against the 36K *M. leprae* protein would be that that T_S cells specifically recognize (an) idiotypic determinant(s) on the 36K responsive T_H cells. Whatever the mechanism of the observed suppression, it is clear the activation of antigen specific suppressor cells places constraints on the use of such proteins for prophylactic immunization. We hope that a suppressor epitope will turn out to be the right explanation, because then the rational design of a synthetic *M. leprae* vaccine would imply the selection of helper epitopes and/or the molecular dissociation of helper from suppressor epitopes²². If the mechanism would appear to be an anti-idiotypic one such a vaccine strategy would not easily result in the circumvention of *M. leprae* reactive suppression in susceptible individuals.

In order to obtain more insight in the specificity and mechanism of the suppression mediated by these T_S clones, we isolated CD4-positive clones from the same patient that proliferated to *M. leprae* by cloning the cells relatively early (96 hours) after a single exposure to *M. leprae*. This procedure was followed because helper T-cell growth was reported to precede that of suppressor T cells in time²³. The thus isolated supposedly T_H clones responded to *M. leprae* specific or cross-reactive determinants, like was observed for tuberculoïd patients and healthy contacts^{16,24} and were HLA-DR or -DP restricted in their response. As shown in Table 2, T_S clones could suppress both

M. leprae specific and cross-reactive T_H clones in a number of cases (1G2, 1D11). Other T_S clones appeared to suppress only some (1E9) or no (1D10) T_H clones, although suppressing the *M. leprae* response of PBMNC. Apart from providing a (negative) control for the specific suppression on T_H clones, the latter observation may either indicate that in order to complete and/or amplify the suppressor circuit, cells other than the T_S clones have to be recruited. Alternatively, the T_S clones may be anti-idiotypic and would not suppress T_H clones (2F9, 2B2) which carry another idiotypic determinant than that seen by the T_S cell. However, probably the most important aspect of the data presented in Table 2 is the fact that the same T_S clones which suppressed T_H responses to cross-reactive mycobacterial antigens (T_S clones 1G2 and 1D11, see also Table 1) are also able

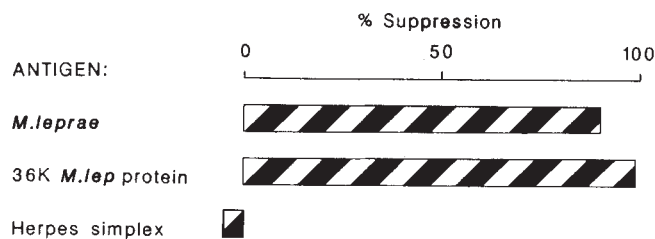


Fig. 2 Antigen-specific suppression of peripheral T-cell responses against a purified *M. leprae* 36K protein. The results are expressed as the percentage suppression after the addition of 3×10^4 T suppressor cells (see legend table 1). The response in the absence of T suppressor cells against the 36K protein was 12,020 c.p.m., against *M. leprae* 27,040 c.p.m., and against HSV 14,255. Responses with no antigen in the culture were 1,030 c.p.m. in the absence and 428 c.p.m. in the presence of the T suppressor cells. Standard deviations did not exceed 20%. The suppressor T-cell line was tested in this experiment. Similar results have now been obtained with T_S clones. The tested 36K protein was purified by affinity chromatography using monoclonal antibody F47-9-1 (refs 14, 15) coupled to cyanogen bromide-activated Sepharose 4B, followed by elution with 0.1 M diethylamine pH 11.5. The protein was tested in a concentration of $5.6 \mu\text{g ml}^{-1}$.

Table 2 T_S clones suppress the *M. leprae* response of some but not all *M. leprae* reactive T_H clones of the same patient

	Proliferative T _H clone		
	2F9	2B2	2F6
Proliferation to <i>M. leprae</i> :	+	+	+
Proliferation to other mycobacteria:	—	±	+
T _S clone added	Percentage suppression		
1G2	41 ¹	47	63
1D11	55	52	62
1E9	5	-1	30
1D10	-51	-13	4

Results are expressed as the percentage suppression (see table 1). *M. leprae* specific, (2F9), partly crossreactive (2B2) or completely crossreactive (2F6) T_H clones of the same patient were induced to proliferate by autologous irradiated APC and *M. leprae* antigen as described recently^{16,26}. Briefly, 10⁴ TLC cells were cocultured with 5 × 10⁴ APC (40 Gy irradiated) and Dharmendra lepromin (1:80) for 72 h, after which ³H-TdR was added as described in Fig. 1 legend. The mean c.p.m. of the T_H clones in the absence of suppressor T cells were 36,780 for 2F9, 8,252 for 2B2 and 8,694 for 2F6. At the start of the culture either no or 3 × 10⁴ suppressor cells were added. Background proliferation in cultures containing T_H and T_S cells without antigen was always less than 440 c.p.m. The percentage of suppression was calculated as described in the legend of Fig. 1. Standard deviations did not exceed 20%.

to suppress a T_H clone directed against an *M. leprae* specific epitope. This might indicate that cross-reactive T_S clones as described in this paper although not *M. leprae* specific, could play an essential role in the *M. leprae* specific unresponsiveness observed in lepromatous leprosy. It is interesting that in several instances a marked depression of *M. tuberculosis* and BCG responses has been noted in lepromatous leprosy patients^{25,26}. In the light of our findings this may be explained at least in part by *M. leprae* induced crossreactive T_S cells. We deliberately selected a borderline lepromatous leprosy patient, which was not unresponsive to *M. leprae* in order to be able to generate both T_H and T_S clones. Therefore we have to be cautious in generalizing this observation because it is generally assumed that lepromatous leprosy patients are nonresponsive to *M. leprae* but remain good responders towards other mycobacteria (reviewed in refs 1, 5). Interestingly, it has recently been suggested that such patients—in contrast to tuberculoid patients and healthy contacts—do not respond to cross reactive or common antigens on other mycobacteria, but rather to the species specific antigens except of course the *M. leprae* specific ones (refs 25, 27, 28; G. A. W. Rook, submitted for publication). The question then arises why the *M. leprae* induced crossreactive T_S cells could not also suppress T_H responses against the species specific antigens of other mycobacteria than *M. leprae*. Extending the antigen bridging model (refs 20, 21) we propose that these T_S cells might only suppress T_H cells when both the suppressor epitope (that is, the crossreactive *M. leprae* suppressor epitope) and the helper epitope to which the *M. leprae* reactive T_H cells were induced originally (that is, either a *M. leprae* specific or a crossreactive helper epitope) are expressed by the same molecule. This prerequisite is not met in the case of T_H responses against the species specific helper epitopes of other mycobacteria than *M. leprae*: since these helper epitopes by definition are not situated on molecules shared with *M. leprae*, T_H responses against these non-*M. leprae* epitopes can evade suppression by cross-reactive T_S cells irrespective of the presence of a crossreactive suppressor epitope on that same molecule. Qualitative and quantitative differences in the sensitization towards those species specific helper epitopes in lepromatous leprosy patients might thus account for at least part of the heterogeneity observed in responsiveness against other mycobacteria in lepromatous leprosy.

Since the observed suppression of peripheral T_H-cell responses and T_H clones might be explained by cytotoxicity against either responding antigen specific T_H cells or APC, we investigated whether peripheral cells activated by *M. leprae* or HSV, or autologous Epstein-Barr virus-B cells, pulsed or unpulsed with *M. leprae*, or *M. leprae* reactive T_H clones could be lysed by suppressor cells. However, no specific lysis of these targets by the suppressor T-cell line and T_S clones was observed, whereas a control cytotoxic T-cell line specific for HLA-A2 strongly lysed all targets tested. Therefore these results excluded cytotoxicity as a possible mechanism of suppression.

Finally, we were interested to know whether these T_S clones were HLA-restricted or not and if so by which HLA molecules and epitopes. Modlin *et al.*²⁹ have recently claimed that suppressor T cells from the skin lesions of lepromatous leprosy patients may be restricted by HLA-DR. Our own preliminary data obtained with both panel and family studies in which the T_S cells were mixed with allogeneic T_H cells have indeed also shown evidence for some kind of restriction. However, this was not simply associated with HLA-class I, -DR or -DQ alleles. Thus far, we have been unable to abolish the suppression with HLA-class I or HLA-DQ specific monoclonal antibodies, but HLA-DR antibodies could do so in certain T_H-T_S cell combinations. Thus, although HLA-DR might be involved, the genetic restriction of T_S cells seems to be more complex than that of T_H cells.

The lines and clones described in this paper can be made available to interested colleagues.

We thank Professor J. J. van Rood for his support for this project; Dr. R. C. Good for *M. leprae* antigen; Professor D. L. Leiker for help and cooperation; the patients for donating blood; Earl Johanns and John Haanen for technical assistance; Fons UytdeHaag for HSV antigens; Madeleine de Wit for isolating the 36K protein; Arend Kolk for monoclonal antibodies; Jos Pool for assistance in the CML test and in preparing serum pools, Annemarie Termijtelen for helpful discussions and Ellen van der Willik for preparing of the manuscript. This study was supported in part by the Foundation for Medical Research (FUNGO grant 900-509-099), the Immunology of Leprosy (IMMLEP) component of the UNDP/WORLD Bank/WHO Special Programme for Research and Training in Tropical Diseases, the Netherlands Leprosy Relief Association (NSL), and the J. A. Cohen Institute for Radiopathology and Radiation Protection (IRS).

Received 21 January; accepted 21 April 1986.

- Bloom, B. R. & Godal, T. *Rev. infect. Dis.* **5**, 765-780 (1983).
- Rees, R. J. W. *Nature* **211**, 576 (1966).
- Patel, P. J. & Lefford, M. J. *Infect. Immun.* **19**, 87-93 (1978).
- Lowe, C., Brett, S. J. & Rees, R. J. W. *Clin. exp. Immun.* **61**, 336-342 (1985).
- Bloom, B. R. & Mehra, V. *Immun. Rev.* **80**, 5-28 (1984).
- Mehra, V., Mason, L. H., Fields, J. P. & Bloom, B. R. *J. Immun.* **123**, 1813-1817 (1979).
- Mehra, V., Brennan, P. J., Rada, E., Convit, J. & Bloom, B. R. *Nature* **308**, 194-196 (1984).
- Nath, I., van Rood, J. J., Mehra, N. K. & Vaidya, M. C. *Clin. exp. Immun.* **42**, 203-210 (1980).
- Sathish, M., Bhutani, L. K., Sharma, A. K. & Nath, I. *Infect. Immun.* **42**, 890-899 (1983).
- Salgame, P. R., Mahadevan, P. R. & Antia, N. H. *Infect. Immun.* **40**, 1119-1126 (1983).
- Blue, M. L., Daley, J. F., Levine, H. & Schlossman, S. F. *J. Immun.* **134**, 2281-2286 (1985).
- Faracet, J. P. *et al. Eur. J. Immun.* **15**, 1067-1073 (1985).
- Young, R. A. *et al. Nature* **316**, 450-452 (1985).
- Klatzer, P. R., de Wit, M. Y. L. & Kolk, A. H. J. *Clin. exp. Immun.* **62**, 468-473 (1985).
- Engers, H. D. *et al. Infect. Immun.* **48**, 603-605 (1985).
- Ottenhoff, T. H. M. *et al. Nature* **319**, 66-68 (1986).
- Wicker, L. S., Katz, M., Sercarz, E. & Miller, A. *Eur. J. Immun.* **14**, 442-447 (1984).
- Goodman, J. W. & Sercarz, E. A. *Rev. Immun.* **1**, 465-498 (1983).
- Lehner, T., Mehler, A., Avery, J., Jones, T. & Caldwell, J. *J. Immun.* **135**, 1437-1442 (1985).
- Oki, A. & Sercarz, E. *J. exp. Med.* **161**, 897-911 (1985).
- Krzych, U., Fowler, A. V. & Sercarz, E. *J. exp. Med.* **162**, 311-323 (1985).
- Mitchison, N. A. *Nature* **308**, 112-113 (1984).
- Bensussan, A., Acuto, O., Hussey, R. E., Milanese, C. & Reinherz, E. L. *Nature* **311**, 565-567 (1984).
- Mustafa, A. S. *et al. Nature* **319**, 63-66 (1986).
- Reitan, L. J., Closs, O. & Beihau, A. *Int. J. Lepr.* **50**, 455-467 (1982).
- Godal, T., Myklestad, B., Samuel, D. R. & Myrvang, B. *Clin. exp. Immun.* **9**, 825-831 (1971).
- Smelt, A. H. M., Rees, R. J. W. & Liew, F. Y. *Clin. exp. Immun.* **44**, 507-511 (1981).
- Stanford, J. L., Nye, P. M., Rook, G. A. W., Samuel, N. M. & Fairbank, A. A. *Lepr. Rev.* **52**, 321-327 (1981).
- Modlin, R. L. *et al. Nature* (in the press).
- Ottenhoff, T. H. M., Elferink, B. G., Hermans, J. & de Vries, R. R. P. *Human Immun.* **13**, 105-116 (1985).
- Elferink, B. G., Ottenhoff, T. H. M. & de Vries, R. R. P. *Scand. J. Immun.* **22**, 585-589 (1985).
- Haanen, J. B. A. G. *et al. Scand. J. Immun.* **23**, 101-108 (1986).

Retinal ganglion cells lose response to laminin with maturation

J. Cohen*, J. F. Burne*, J. Winter† & P. Bartlett‡

*MRC Neuroimmunology Project, Department of Zoology, University College London, London WC1E 6BT, UK

†Sandoz Institute for Medical Research, 5, Gower Place, London WC1E 6BN, UK

‡Walter and Eliza Hall Institute of Medical Research, Royal Melbourne Hospital, Victoria 3050, Australia

The decisive role played by adhesive interactions between neuronal processes and the culture substrate in determining the form and extent of neurite outgrowth *in vitro*^{1,2} has greatly influenced ideas about the mechanisms of axonal growth and guidance in the vertebrate nervous system. These studies have also helped to identify adhesive molecules that might be involved in guiding axonal growth *in vivo*. One candidate molecule is laminin, a major glycoprotein of basal laminae³ which has been shown to induce a wide variety of embryonic neurones to extend neurites in culture⁴⁻⁸. Moreover, laminin is found in large amounts in injured nerves that can successfully regenerate but is absent from nerves where regeneration fails⁹⁻¹¹. However, it is unclear to what extent the mechanisms that regulate axonal regeneration also operate in the embryo when axon outgrowth is initiated. Here we have examined the substrate requirements for neurite outgrowth *in vitro* by chick embryo retinal ganglion cells, the only cells in the retina to send axons to the brain. We show that while retinal ganglion cells from embryonic day 6 (E6) chicks extend profuse neurites on laminin, those from E11 do not, although they retain the ability to extend neurites on astrocytes via a laminin-independent mechanism. This represents the first evidence that central nervous system neurones may undergo a change in their substrate requirements for neurite outgrowth as they mature.

In the chick retina most retinal ganglion cells (RGC) become post-mitotic between E3 and E7 (ref. 12), and during this period axon outgrowth is largely confined within the retina, optic stalk and tract. By E11 the production of RGC has virtually ceased and the majority of optic axons have entered their target tissue, the optic tectum¹³. To compare the behaviour of RGC in cultures of E6 and E11 retinal cells, we have used a monoclonal antibody against chicken Thy-1 glycoprotein¹⁴, an immunologically distinct homologue of mammalian Thy-1 (ref. 15). As previously shown for the immature rat retina^{16,17}, Thy-1 is restricted to RGC in cryostat sections of the embryonic chick retina (Fig. 1). Whereas optic nerve transection experiments in the rat suggest

that some bipolar and amacrine neurones might also express Thy-1 in the mature retina¹⁸, most of these cells differentiate after E11 in the chick and are unlikely to account for a significant proportion of the Thy-1⁺ cells seen in our cultures.

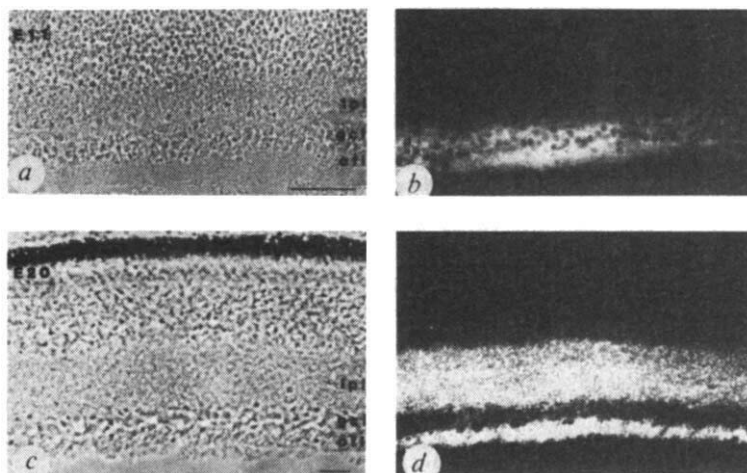
Immunoreactive Thy-1 antigen was first detectable *in situ* at E5 as faint punctate deposits in the immature RGC layer (not shown). By E11 the antigen was also detectable in the optic fibre layer (OFL) (Fig. 1a, b), and at E20 it was found in regions containing RGC cell bodies, their axons and dendrites, but was absent from other retinal layers (Fig. 1c, d). In retinal suspensions at E6 and E11, 2-5% of these cells were Thy-1⁺ (not shown), including more than 90% of the larger (>10 µm diameter) cell bodies.

When cells dissociated from E6 or E11 retina were plated on polylysine (PL) coated glass coverslips at a sufficiently low density ($\leq 50,000$ cells cm⁻²) to keep them separate, >80% of the Thy-1⁺ cells attached within 2 h. However, after 16 h <10% had extended neurites and most had begun to degenerate. Similar results were obtained with a variety of substrates, including polyornithine, fibronectin, rat-tail collagen and fibroblast ghosts (not shown). However, when E6 cells were grown on laminin, most Thy-1⁺ cells (>75%; Fig. 3) extended long neurites (Fig. 2a, b); profuse, non-fasciculated outgrowth of Thy-1⁺ processes, which also stained with a monoclonal anti-neurofilament antibody RT97 (not shown), was evident within a few hours of plating, being maximal at 48-72 h, by which time some Thy-1⁺ RT97⁺ neurites extended distances in excess of 50 cell diameters (>600 µm). By contrast, <10% of Thy-1⁺ cells in cultures of E11 retina extended neurites on laminin and this growth pattern was indistinguishable from that on polylysine (Figs 2c, d, 3). The same stage-dependent response was also seen on a substrate (ACM) consisting of polylysine binding material from culture medium conditioned by rat type-1 astrocytes¹⁹ (Fig. 3). Although we have not characterized the factor in ACM, previous studies with a variety of culture medium-derived neurite-promoting factors make it likely that the activity of ACM resides in a complex of laminin with heparan sulphate proteoglycan^{20,21}.

A possible explanation for the failure of E11 Thy-1⁺ neurones to extend neurites *in vitro* is that as RGC mature, they become intrinsically unable to regenerate their axons. However, previous studies have demonstrated that rat RGC, which had extended axons to their retino-receptive targets in the brain, are able to regrow neurites when cultured on some substrates^{22,23}. Moreover, after optic nerve transection in adult rats, RGC have been shown to regrow axons into sciatic nerve grafts²⁴. These observations on mammalian RGC suggested that E11 chick RGC could extend neurites *in vitro* if given the appropriate

Fig. 1 Localization of Thy-1 antigen in embryonic chick retina. Phase-contrast (a, c) and fluorescence (b, d) photomicrographs of E11 (a, b) and E20 (c, d) cryostat sections of retina. b, At stage E11, monoclonal anti-Thy-1 antibody binds only to cells in the retinal ganglion cell layer (RGC) and axons in the optic fibre layer (OFL) (d). At E20, immunofluorescence staining is more intense and extends also to the inner plexiform layer (IPL) but is absent from all other retinal layers. Both a and c show dorsal-central regions of retina adjacent to the optic nerve head. Scale bars, 50 µm.

Methods. Whole eyes were fixed in paraformaldehyde-lysine-periodate²⁵, cryoprotected with sucrose, infiltrated with OCT embedding compound, and frozen in liquid nitrogen. Cryostat sections (6-10 µm) were cut in the horizontal plane, dried on polylysine-coated (PL) slides and incubated overnight at 4°C with monoclonal anti-chick Thy-1 antibody (ascites fluid diluted 1:100) followed successively by incubation for 2 h at room temperature in biotinylated sheep anti-mouse immunoglobulin and streptavidin-fluorescein (Amersham; diluted 1:100). Sections were mounted in glycerol containing 1,4-diazobicyclo-(2,2,2)octane (DABCO) to inhibit fading of fluorescence²⁷, and photographed on a Zeiss Universal microscope.



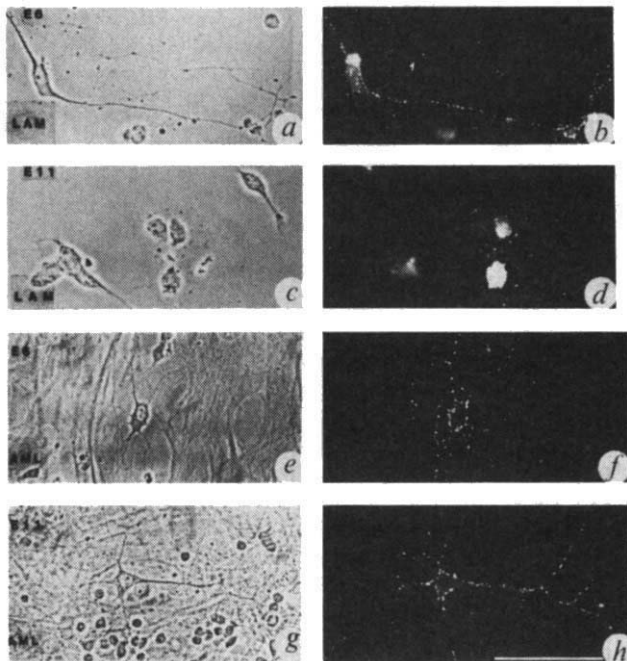


Fig. 2 Effect of maturation on neurite extension by Thy-1⁺ retinal neurones grown on laminin and astrocytes. The cells were photographed using phase-contrast (a, c, e, g) or fluorescein (b, d, f, h) optics. Note that while E6 Thy-1⁺ neurones extend neurites on laminin (LAM) (a, b) and type-1 astrocytes (AML) (e, f), E11 Thy-1⁺ neurones fail to extend neurites on LAM (c, d) but do so on AML (g, h). All Thy-1⁺ process-bearing cells also stained with a monoclonal antibody (RT97) against neurofilaments (not shown). On AML (e-h) Thy-1⁺ cells assumed a more complex morphology than on laminin (a-d). Scale bar, 50 μ m.

Methods. Retinas were separated from choroidal and scleral tissue, dissected free of pigment epithelium and incubated at 37 °C for 20 min in Ca²⁺, Mg²⁺-free Earle's balanced salts (BSS) plus 0.1% trypsin and bovine serum albumin (BSA, 3 mg ml⁻¹; Sigma). Tissue was removed and resuspended in BSS plus soybean trypsin inhibitor (SBTI, 0.52 mg ml⁻¹; Sigma), DNase (0.04 mg ml⁻¹; Sigma) and BSA and dissociated by repetitive trituration in a wide-bore Pasteur pipette to give a single-cell suspension. Homogeneous cultures of rat type-1 astrocytes (AML) were prepared from neonatal cerebral cortex tissue as described previously²⁸. (Rat, rather than chick astrocytes were used as a growth substrate as immature chick astrocytes do not express the astrocyte-specific marker GFAP²⁸ and could not, therefore, be reliably monitored for their purity *in vitro*. Preliminary experiments established that the neurite-outgrowth-promoting activity of rat type-1 astrocytes was as effective with chick neurones as previously shown for rat neurones²⁵.) Laminin substrate was prepared by incubating PL-coated glass coverslips at room temperature for 2 h in a solution of laminin (10 μ g ml⁻¹; BRL) in serum-free culture medium (see below), followed by two brief rinses in medium. Retinal cells (50,000) were plated onto pre-coated glass coverslips in Falcon multi-well plates and grown in Ham's F12 medium (Flow) supplemented with 0.5% fetal calf serum and the following (μ g ml⁻¹): 100 BSA, 5 insulin, 100 transferrin, 6 progesterone, 16.1 putrescine and 5.6 selenium, as modified from Bottenstein and Sato³⁹. All cultures were incubated for 48 h at 37 °C in a humidified atmosphere of 5% CO₂, 95% air. Living cultures were labelled with monoclonal anti-Thy-1 antibody (ascites fluid diluted 1:100) for 30 min at room temperature, followed by biotinylated anti-mouse immunoglobulin and fluorescein-streptavidin as described in Fig. 1 legend, except that incubation times were shortened to 30 min. Coverslips were fixed in 95% ethanol/5% acetic acid, mounted in glycerol/DABCO and labelled cells photographed as described in Fig. 1 legend.

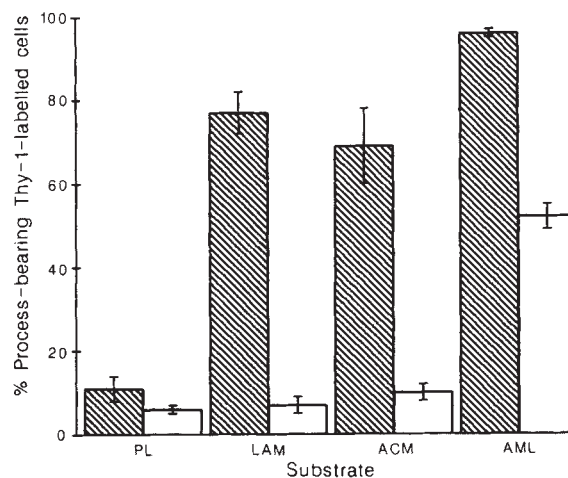


Fig. 3 Effect of laminin (LAM), ACM and AML on neurite outgrowth by E6 (■) and E11 (□) Thy-1⁺ neurones. Most of E6 Thy-1⁺ neurones extended neurites on LAM, ACM and AML while only AML induced neurite outgrowth by E11 Thy-1⁺ neurones. Retinal cells (50,000) from E6 or E11 retinas were plated onto glass coverslips pre-coated with PL, LAM or AML as described in Fig. 2 legend. In experiments using ACM as substrate (PL-binding material from astrocyte-conditioned medium), PL-coated coverslips were incubated overnight in culture medium conditioned over confluent monolayers of rat astrocytes for at least 24 h. After 48 h of culture, coverslips were labelled with anti-Thy-1 antibody as described in Fig. 2 legend. After fixation the cultures were co-labelled with RT97 (ascites fluid, 1:200) for 30 min at room temperature followed by rhodamine-conjugated anti-mouse immunoglobulin 1:100; Cappel). Coverslips were then washed and mounted as described in Fig. 2 legend. The total number of Thy-1⁺ cells per coverslip (≥ 50 cells), and the number with Thy-1⁺/RT97⁺ processes longer than 3 cell diameters, were counted. The results shown are the mean ($\pm$ s.e.m.) of at least four separate experiments. At E6 a significant difference (Student's *t*-test) was found between PL and LAM ($P < 0.0006$), PL and ACM ($P < 0.04$) and between PL and AML ($P < 0.0001$). At E11 a significant difference was found between PL and AML ($P < 0.0004$) but no significant difference was found between PL and LAM ($P > 0.7$) or between PL and ACM ($P > 0.2$).

substrate. We therefore cultured retinal cells on monolayers of purified rat brain astrocytes (AML) which consisted of >95% glial fibrillary acidic protein-positive (GFAP⁺) type-1 astrocytes with a flat fibroblast-like morphology¹⁹. Such cells are known^{25,26} to provide a highly adhesive substrate for the growth of neurites *in vitro* and possibly also *in vivo*^{27,28}. We found that on AML the majority of Thy-1⁺ neurones in cultures of both E6 and E11 retina extended long neurites (Figs 2e-h, 3). Moreover, the neurite-promoting effect of AML on Thy-1⁺ neurones seemed to depend on direct contact with the astrocytes and not on astrocyte-derived soluble factors, as the effect was abolished in co-cultures arranged to prevent contact²⁵.

An alternative explanation for our findings is that during maturation, RGC surface receptors for laminin are lost or modified. However, evidence against receptor down-regulation is provided by our experiments (not shown) with a monoclonal antibody, JG22, which is specific for a cell surface receptor complex that binds both laminin and fibronectin^{29,30}. When trypsinized suspensions of E6 and E11 retina were allowed to settle onto coverslips and labelled with JG22 antibody, over 80% of cells in the larger size range (>10 μ m diameter), corresponding to the majority of Thy-1⁺ cells on duplicate coverslips, were JG22⁺ at both ages, and labelling was more intense at E11 than at E6. When cultured on a substrate permissive for neurite outgrowth, RGC cell bodies, neurites and growth cones were JG22⁺. Also, in immunostained tissue sections taken from the retinas of E6 and E11 chicks (not shown) RGC axons in the optic fibre layer and the optic nerve head were intensely JG22⁺.

along their entire length. Therefore, these findings raise the possibility that the receptor complex on E11 RGC is modified so as to reduce its affinity for laminin. Alternatively, transduction mechanisms within the cells may be uncoupled from the binding event.

What is the nature of the molecules on the astrocyte surface which promote neurite outgrowth? Whereas brain astrocytes can produce laminin *in vitro*³¹, the failure of either pure laminin or ACM substrate to evoke a response from E11 RGC argues that the neurite-promoting activity of the astrocyte cell surface resides in either a modified form of laminin not represented in ACM, or in an entirely different molecule. Evidence against the first possibility was provided by further experiments (not shown) with JG22 antibody: whereas the antibody blocked most neurite outgrowth by E6 RGC on either laminin or ACM substrate, it failed to inhibit outgrowth by E6 or E11 RGC on AML. It may be that the neurite-promoting molecule on the astrocyte surface is one of the cell adhesion molecules previously shown to influence neurone-astrocyte interactions^{32,33}.

From days 6 to 11 of incubation in the chick embryo, maturation of the optic pathway is characterized by the transition from a predominantly retinal to a tectal phase of RGC axon outgrowth¹³. It is possible, therefore, that the change in substrate requirements for growth *in vitro* reflects a response by RGC *in vivo* to changes in the composition of the substrate that their growth cones encounter as they traverse the optic pathway. Such changes could be regulated by temporal or spatial mechanisms. Evidence for the former has come from an immunohistochemical survey with anti-laminin antibodies of the embryonic chick optic pathway throughout development. We have shown, using a sensitive immunohistochemical method, that in addition to its presence in basement membranes at the inner limiting membrane and pigment epithelium³⁴, laminin is also expressed by neuroepithelial cells throughout the optic pathway, within the retina, optic fissure, stalk and tectum during the earliest stages of maturation (between E3 and E7), becoming restricted to basement membrane locations only after E8 (ref. 35 and J.C., J.F.B. and J.W., unpublished observations). These findings suggest that laminin is a component of the physiological substrate for RGC axon extension during the earliest phase of outgrowth *in vivo*, and provide an explanation for our observations of the response of E6 RGC to growth on laminin *in vitro*. The mechanism whereby the maturation of RGC is accompanied by a loss of response to laminin remains to be elucidated. The fact that the timing of this change coincides with the arrival of the bulk of optic axons at the tectum raises the possibility that an alteration in the substrate requirements for axon extension may be one of the consequences of the transition from a target-independent to a target-dependent phase of RGC maturation. Future experiments should allow us to test whether optic tectum-derived factors can influence the substrate requirements *in vitro* for neurite outgrowth by RGC as a function of maturation.

We thank D. Gottlieb for supplying JG22 antibody, John Wood for RT97 antibody and Martin Raff and other members of the Neuroimmunology Group for helpful suggestions.

Received 18 March; accepted 21 May 1986.

- Bray, D. *Nature* **244**, 93-96 (1973).
- Letourneau, P. C. *Dev Biol.* **44**, 92-101 (1975).
- Timpl, R. H. *et al.* *J. biol. Chem.* **254**, 9933-9937 (1979).
- Baron von Evercooren *et al.* *J. Neurosci. Res.* **8**, 179-193 (1982).
- Manthorpe, M. *et al.* *J. Cell Biol.* **97**, 1882-1890 (1983).
- Rogers, S. L. *et al.* *Dev Biol.* **98**, 212-220 (1983).
- Smallheiser, N. R., Crain, S. M. & Reid, L. M. *Dev Brain Res.* **12**, 136-140 (1984).
- Calof, A. L. & Reichardt, L. F. *Dev Biol.* **106**, 194-210 (1984).
- Bignami, A., Chi, N. H. & Dahl, D. *Expl Neurol.* **85**, 226-236 (1984).
- Hopkins, J. M., Ford-Holevinski, T. S., McCoy, J. P. & Agranoff, B. W. *J. Neurosci.* **5**, 3030-3038 (1985).
- Liesi, P. *EMBO J.* **4**, 2505-2511 (1985).
- Kahn, A. J. *Brain Res.* **63**, 285-290 (1973).
- Rager, G. H. *Adv. Anat. Embryol. Cell Biol.* **63**, 1-92 (1980).
- Sinclair, C. M., Bartlett, P. F., Grieg, D. I. & Jeffrey, P. L. *Brain Res.* (in the press).
- Rostas, J. A. P., Sherman, A., Sinclair, C. M. & Jeffrey, P. L. *Biochem. J.* **213**, 143-152 (1983).
- Beale, R. & Osborne, N. N. *Neurochem. Int.* **4**, 587-595 (1982).
- Barnstable, C. J. & Drager, H. C. *Neuroscience* **11**, 847-855 (1984).

- Perry, V. H., Morris, R. J. & Raisman, G. *J. Neurocytol.* **13**, 809-824 (1984).
- Raff, M. C. *et al.* *J. Neurosci.* **3**, 1289-1300 (1983).
- Lander, A. D., Fujii, D. K. & Reichardt, L. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2183-2187 (1985).
- Davis, G. E., Manthorpe, M., Engvall, E. & Varon, S. *J. Neurosci.* **5**, 2662-2671 (1985).
- Leifer, D., Lipton, S. A., Barnstable, C. J. & Masland, R. H. *Science* **224**, 303-306 (1984).
- McCaffery, C. A., Raju, T. R. & Bennett, M. R. *Dev Biol.* **104**, 441-448 (1984).
- So, K.-F. & Aquayo, A. J. *Brain Res.* **32**, 349-354 (1985).
- Noble, M., Fok-Seang, J. & Cohen, J. *J. Neurosci.* **4**, 1892-1903 (1984).
- Fallon, J. S. *J. Cell Biol.* **100**, 198-207 (1985).
- Silver, J. & Sapiro, J. *J. comp. Neurol.* **202**, 521-538 (1981).
- Kryanek, S. & Goldberg, S. *Dev Biol.* **84**, 41-50 (1981).
- Greve, J. M. & Gottlieb, D. L. *J. cell. Biochem.* **18**, 221-230 (1982).
- Horwitz, A. *et al.* *J. Cell Biol.* **101**, 2134-2144 (1985).
- Liesi, P., Dahl, D. & Vaheri, A. *J. Cell Biol.* **96**, 920-924 (1983).
- Grummet, M., Hoffman, S. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 267-271 (1984).
- Keilhauer, G., Faissner, A. & Schachner, M. *Nature* **316**, 728-730 (1985).
- Adler, R., Jordan, J. & Hewitt, A. T. *Dev Biol.* **112**, 100-114 (1985).
- Cohen, J. *Neurosci. Lett. Suppl.* **22**, 36 (1985).
- McLean, I. W. & Nakane, P. K. *J. Histochem. Cytochem.* **22**, 1077-1083 (1984).
- Johnson, G. D. *et al.* *J. immun. Meth.* **55**, 231-242 (1982).
- Lemmon, V. *Dev Brain Res.* **23**, 111-120 (1985).
- Bottenstein, J. & Sato, G. H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 515-517 (1979).

Chloride and potassium channels in cystic fibrosis airway epithelia

Michael J. Welsh & Carole M. Liedtke

Laboratory of Epithelial Transport and Pulmonary Division, Department of Internal Medicine, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA and the Cystic Fibrosis Center, Department of Pediatrics, Developmental Genetics and Anatomy, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106, USA

Cystic fibrosis, the most common lethal genetic disease in Caucasians, is characterized by a decreased permeability in sweat gland duct and airway epithelia. In sweat duct epithelium, a decreased Cl^- permeability accounts for the abnormally increased salt content of sweat¹. In airway epithelia a decreased Cl^- permeability, and possibly increased sodium absorption, may account for the abnormal respiratory tract fluid^{2,3}. The Cl^- impermeability has been localized to the apical membrane of cystic fibrosis airway epithelial cells⁴. The finding that hormonally regulated Cl^- channels make the apical membrane Cl^- permeable in normal airway epithelial cells⁵ suggested abnormal Cl^- channel function in cystic fibrosis. Here we report that excised, cell-free patches of membrane from cystic fibrosis epithelial cells contain Cl^- channels that have the same conductive properties as Cl^- channels from normal cells. However, Cl^- channels from cystic fibrosis cells did not open when they were attached to the cell. These findings suggest defective regulation of Cl^- channels in cystic fibrosis epithelia; to begin to address this issue, we performed two studies. First, we found that isoprenaline, which stimulates Cl^- secretion, increases cellular levels of cyclic AMP in a similar manner in cystic fibrosis and non-cystic fibrosis epithelial cells. Second, we show that adrenergic agonists open calcium-activated potassium channels, indirectly suggesting that calcium-dependent stimulus-response coupling is intact in cystic fibrosis. These data suggest defective regulation of Cl^- channels at a site distal to cAMP accumulation.

Figure 1 shows a model of the cellular mechanism of Cl^- secretion by airway epithelia⁶. Chloride enters the cell at the basolateral membrane coupled to Na^+ on an electrically neutral co-transport process. Chloride then leaves the cell across a Cl^- -conductive apical membrane; regulation of the Cl^- conductance determines, in part, the rate of transepithelial Cl^- secretion. Chloride channels produce the apical Cl^- conductance. Using the patch-clamp technique⁷, we recently described a Cl^- channel from the apical membrane of normal human airway epithelial cells which had properties different from those of previously described Cl^- channels^{5,8}. The ion selectivity, response to inhibitors and activation by secretagogues indicate that this Cl^-

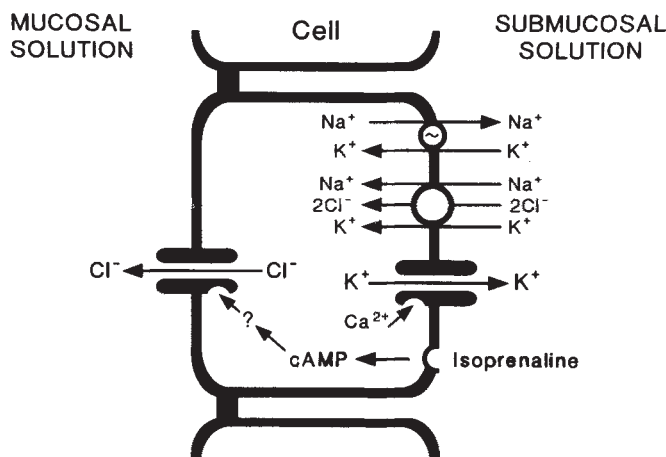


Fig. 1 Model of the cellular mechanism of Cl^- secretion. Chloride enters at the basolateral membrane via a NaKCl co-transport process that is inhibited by loop diuretics. The energy to drive intracellular Cl^- accumulation above electrochemical equilibrium comes from the movement of Na^+ down its electrochemical gradient. A low intracellular Na^+ concentration is maintained by the ouabain-inhibitable $(\text{Na}^+ + \text{K}^+)\text{ATPase}$. Potassium that enters the cell in exchange for Na^+ and coupled to Cl^- , exits from the cell across the basolateral membrane through Ca^{2+} -regulated K^+ channels. Chloride leaves the cell across the apical membrane through Cl^- channels.

channel is responsible for the apical Cl^- conductance in airway epithelia.

The decreased Cl^- permeability in cystic fibrosis (CF) apical membranes is therefore most probably due to abnormal Cl^- channel function. There are several possible ways to explain such a defect: it might result from an absence of Cl^- channels (such as faulty insertion into the cell membrane), from the presence of Cl^- channels with abnormal conductive or ion-selective properties, from inhibition of the Cl^- channels, or from faulty regulation of apical membrane Cl^- channels.

To address these possibilities, we examined single-channel currents in primary cultures of CF airway epithelial cells, to look for Cl^- channels with the same properties as those observed in non-CF cells. In excised cell-free patches of membrane from four preparations, we observed single-channel currents that were characteristic of Cl^- channels. Figure 2a shows tracings of the

current from an excised Cl^- channel, and Fig. 2b shows the single-channel current voltage ($I-V$) relation of that same channel. Figure 2b also shows that cation substitutions did not alter the $I-V$ relation. In six patches the channel was exposed to cation gradients, either Na^+ or K^+ , without a shift in reversal potential. In contrast, Fig. 2c shows that a Cl^- concentration gradient shifted the reversal potential, as expected for a channel that is selective for Cl^- over Na^+ . Calculation of the relative Cl^- to Na^+ permeability, $P(\text{Cl}/\text{Na})$, from the Goldman-Hodgkin-Katz equation yielded a value of 6.5 ± 0.4 (mean $\pm$ s.e.m., $n = 3$). The $I-V$ relation of a Cl^- channel from a non-CF cell is shown in Fig. 2b for comparison.

Several features indicate that in excised patches these channels are identical to non-CF Cl^- channels that are responsible for the apical Cl^- conductance and Cl^- secretion⁵. First, the conductance of the channel (26 ± 2 pS measured at 0 mV with

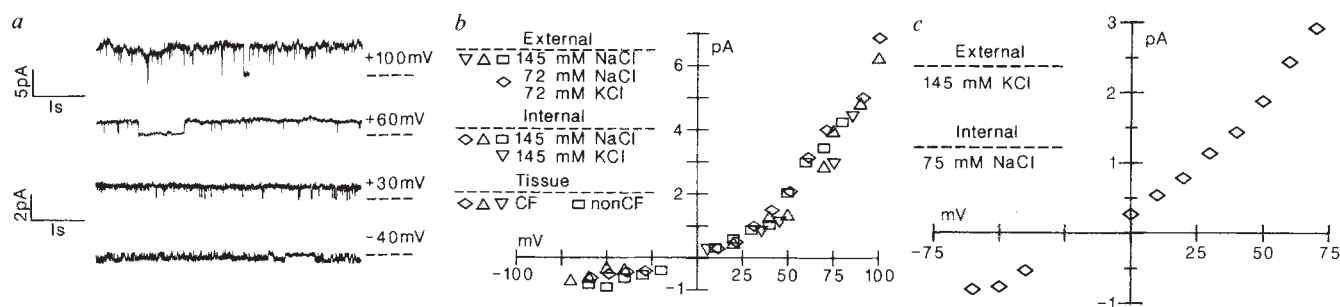


Fig. 2 Tracings and current-voltage relations of Cl^- channels. **a**, Single Cl^- channel currents recorded from an excised, inside-out patch at four holding voltages. Outward currents are shown as upward deflections. The current level when the channel is closed is given by the dashed line. The tracing shows substantial open-channel noise, a characteristic of the Cl^- channel. This channel was open most of the time and did not show marked voltage dependence of kinetics. The external (pipette) solution contained 72 mM NaCl , 72 mM KCl and 1 mM Ca^{2+} and the internal (bath) solution contained 145 mM NaCl and 5 mM EGTA (estimated Ca^{2+} concentration < 1 nM). **b**, Current-voltage relation obtained from the channel shown in Fig. 2a ($\diamond$); $I-V$ relation with symmetrical 145 mM NaCl solutions (Δ); and the relation when the external solution was 145 mM NaCl and internal solution was 145 mM KCl (∇). For comparison the $I-V$ relation of a Cl^- channel obtained from non-CF cells is shown ($\square$)⁵. There is some variability in the absolute conductance of Cl^- channels from non-CF cells; however, Cl^- channels from CF cells did not show any appreciable difference from non-CF channels. **c**, $I-V$ relation with 145 mM KCl externally and 75 mM NaCl , 150 mM sucrose and 2.5 mM EGTA internally. The observed reversal potential, approximately -13 mV, is close to the value expected for the Cl^- activity gradient (-15.4 mV). In contrast, the cation gradients across the channel would predict either positive or very large negative values for K^+ - or Na^+ -selective channels, respectively. We identified this channel in 19 of 198 successfully excised patches from CF cells, whereas under similar conditions it was observed in 2–10% of successfully excised patches from non-CF cells⁵.

Methods. The technique used for constructing pipettes and seal formation^{5,12} is similar to that described by Hamill *et al.*⁷. Either a Dagan Model 8900 or a List EPC 7 amplifier was used for voltage clamping and current amplification. For some studies, currents were filtered (500–750 Hz) by an 8-pole Bessel filter, viewed on an oscilloscope and recorded on a strip-chart recorder. The 90% rise-time for a 2-cm deflection on the recorder was 3.5 ms. Results were analysed by hand. For other studies currents were filtered at 1 kHz and recorded (500 μ s sampling rate) and analysed with a laboratory computer system (Indec Systems). Pipette resistance was 2.5–10 M Ω and seal resistance was 3–30 G Ω . During seal formation and when recording in the cell-attached mode, bath solution contained (in mM): 140 NaCl , 1.2 CaCl_2 , 1.2 MgCl_2 and 10 HEPES buffered with KOH (5 mM). All bath and pipette solutions contained 10 mM HEPES buffered to pH 7.2, except the standard bath solution at pH 7.4. Junction potentials that developed in the presence of asymmetrical salt solutions were directly measured with a 3 M KCl electrode and corrected for. Experiments were performed at room temperature (21–23 $^{\circ}\text{C}$). Cells were studied 4 h to 3 days after plating. Voltages are reported in reference to the external surface of the membrane and outward (+) current refers to the flow of cations from internal to external surface of the patch, or anion flow in the opposite direction.

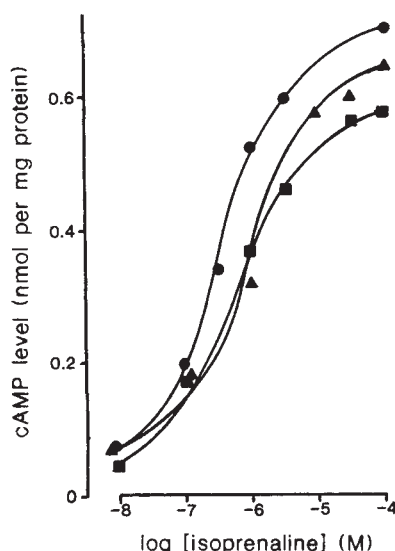


Fig. 3 Effect of isoprenaline on cellular levels of cAMP. Source of cells: ●, CF tracheal epithelium; ■, CF nasal epithelium; ▲, non-CF tracheal epithelium. To measure cellular concentrations of cAMP, the culture medium was replaced with Hank's balanced saline solution (HBSS) containing 10 mM HEPES pH 7.4 and (–)isoprenaline at varying concentrations. Isoprenaline was prepared and diluted in the same buffered salt solution. After incubation for 10 min at 37 °C, the medium was removed and cellular cAMP was released using dilute HCl (0.1 M). Cyclic AMP radioimmunoassay kits (Amersham) were used for measurement of cAMP levels. The cells were treated with 1 M NaOH to dissolve cellular protein which was measured according to the method of Lowry *et al.*¹³, using bovine serum albumin as the standard. Cyclic AMP levels were calculated as nmol cAMP per mg protein.

symmetrical 145 mM Cl^- , $n=10$) is the same as in normal humans. Second, the strong rectification of the $I-V$ relation, even with symmetrical bathing solutions, is quite characteristic. Third, the selectivity for Cl^- over cations is similar. Fourth, there is an appreciable amount of open-channel noise (Fig. 2a) which we find characteristic of Cl^- channels from normal human cells. Fifth, the Cl^- channel blocker diphenylamine 2-carboxylate (1 mM) decreased single-channel current to 53% of control values (at +60 mV, $n=1$, data not shown), a response similar to that found in non-CF Cl^- channels. Sixth, the channel in CF and non-CF cells was not strongly voltage-gated, nor was it acutely regulated by internal Ca^{2+} concentration under these conditions; there was no apparent effect on excised channels of changing from 1 mM internal Ca^{2+} to less than 1 nM Ca^{2+} (5 mM EGTA and nominal Ca^{2+}).

Although excised patches from CF cells had Cl^- channels with conductive and kinetic properties identical to those observed in Cl^- channels from non-CF cells, we never observed a Cl^- channel in the cell-attached recording mode, even though we looked for its activity following each successful seal and treated the cells with isoprenaline (5 μM), adrenaline (1 μM), 8-bromo-cAMP (100 μM) and/or forskolin (100 μM). In non-CF tracheal epithelial cells⁵ the Cl^- channel was observed in the cell-attached mode in 58% of 40 patches that contained Cl^- channels. In contrast, in CF cells the Cl^- channel was not observed cell-attached in the 19 patches that were shown to contain Cl^- channels after excision ($P<0.001$ by χ^2 analysis). These findings suggest that Cl^- impermeability of the apical membrane in CF does not result from the complete absence of Cl^- channels or from the presence of Cl^- channels with abnormal conductive properties. Rather, they suggest that abnormal regulation of Cl^- channels produces the decreased Cl^- permeability⁸.

To investigate regulation of ion transport in CF epithelial cells, two studies were performed. First, we measured intra-

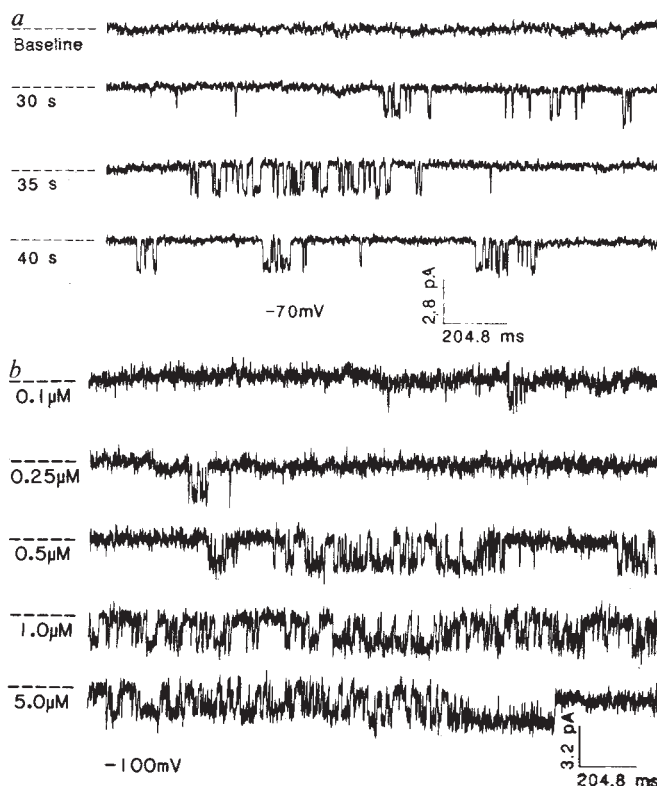


Fig. 4 Potassium channels in CF cells. *a*, Effect of adrenaline on K^+ -channel activity in a cell-attached patch. Sweeps were taken under baseline conditions and at the times indicated, following addition of adrenaline (1 μM) to the bathing solution. The holding potential was -70 mV (membrane hyperpolarized). The pipette solution contained 72 mM KCl, 76 mM NaCl and 1 mM CaCl_2 . Downward deflections represent inward current. The dashed line indicates the current level when all channels are closed. *b*, Effect of internal Ca^{2+} on K^+ -channel activity in an excised, inside-out patch. Tracings were obtained from a single patch held at -100 mV. The external (pipette) solution was 200 mM potassium glutamate and 5 mM MgCl_2 ; internal solution was 140 mM KCl. Internal Ca^{2+} concentrations indicated were achieved by buffering 5 mM EGTA with Ca^{2+} assuming a 1.5×10^{-7} M dissociation constant. **Methods.** Tracheas for cell culture were obtained from three patients at autopsy and nasal polyps were obtained from three patients at the time of polypectomy. Measurement of cAMP was made in cells from one trachea and nasal polyps from three patients, and electrophysiological measurements were made in cells from three tracheas and one culture of nasal polyps. The tissue was washed extensively with sterile HBSS containing (per ml) 100 U penicillin, 100 μg streptomycin, 50 μg tobramycin and 200 mg piperacillin, then incubated at 37 °C with antibiotics for 60 min. The wash and incubation cycles were carried out four times. The tissue was then incubated overnight with HBSS at 7 °C and the epithelium dissected free from the submucosal tissue the next morning. Epithelial cells were isolated according to the method of Liedtke and Tandler¹⁴ with the following modification. Dissected surface epithelium was incubated with collagenase (135 U ml^{-1}), 2 mM EGTA, 2 mM Mg^{2+} and DNase (2 mg ml^{-1}). Dispersed epithelial cells were recovered by centrifugation, washed in sterile HBSS to remove chelators and enzymes, and suspended in culture medium consisting of a 1:1 (v/v) mixture of Ham's F12 and Dulbecco's modified Eagle's medium containing antibiotics as indicated above. Total cell recovery ranged from 38×10^6 to 63×10^6 cells. Cell viability was 95–98% as assessed by fluorescent staining with acridine orange and ethidium bromide. An aliquot of the cell suspension was flown to M.J.W. for studies on single ion channels. The remaining cells were diluted in culture medium supplemented with 5 μM hydrocortisone, 5% fetal calf serum and (per ml) 1 μg insulin, 1 μg transferrin, 4 μg fibronectin and 1 μg epidermal growth factor. For cAMP assays, cells were plated onto collagen-coated plastic culture dishes at 500,000 cm^{-2} . For electrophysiological studies, cells were seeded at 5,000–20,000 cm^{-2} . Cultures were maintained at 37 °C in a humidified 5% CO_2 atmosphere.

cellular levels of cAMP. Although the intracellular factors that directly regulate the Cl^- channel are unknown, circumstantial evidence suggests that cAMP is involved: many secretagogues, including β -adrenergic agonists, increase intracellular levels of cAMP and exogenous addition of cAMP stimulates Cl^- secretion^{9,10}. Figure 3 shows that the β -adrenergic agonist (-)isoprenaline caused a similar dose-dependent increase in cAMP levels in cultured CF and non-CF epithelial cells. The concentration of isoprenaline required to produce a half-maximal increase in cAMP (EC_{50}) was $0.4 \pm 0.1 \mu\text{M}$ in non-CF tracheal epithelium ($n = 3$), $0.5 \pm 0.1 \mu\text{M}$ in CF nasal epithelium ($n = 3$) and $0.4 \mu\text{M}$ in CF tracheal epithelium ($n = 1$). The increase in cAMP was blocked by ($\pm$)propranolol, a β -adrenergic antagonist, indicating that isoprenaline interacts with β -adrenergic receptors (data not shown). Both CF and non-CF cells showed a similar rank-order of potency for adrenergic agonists: isoprenaline > adrenaline > noradrenaline. Similar results have been obtained in the secretory coil of CF and non-CF sweat glands: isoprenaline produced identical increases in cellular levels of cAMP¹¹. Thus, if cAMP regulates the Cl^- channel, the defect in CF would probably lie somewhere distal to the production of cAMP.

Second, to further examine stimulus-response coupling in airway epithelia, we studied the K^+ channels responsible for the basolateral K^+ conductance¹². Stimulation of Cl^- secretion activates both apical Cl^- channels and basolateral K^+ channels. Activation of K^+ channels results from an increase in intracellular Ca^{2+} concentration. CF cells contained K^+ channels with the same conductive and ion-selective properties as cells obtained from non-CF humans and dogs. Figure 4a shows that addition of adrenaline to the bathing solution activated a cell-attached K^+ channel. A similar response was observed on three occasions using isoprenaline and in cells from three of the four preparations studied. To show that the channel was regulated by the internal Ca^{2+} concentration, we used inside-out, cell-free patches as shown in Fig. 4b. Increasing the concentration of Ca^{2+} bathing the cytosolic surface of the patch increased the probability of finding the K^+ channel open. The same response was observed in cells from all four patients. This apparently normal regulation of the Ca^{2+} -activated K^+ channel suggests that the metabolic steps between hormone binding and regulation of the K^+ channel (that is, regulation of intracellular Ca^{2+} concentration) are intact in CF.

In conclusion, these results suggest that CF epithelial cells contain the same apical Cl^- channel as non-CF cells. However, the channel does not open when attached to the cell. Thus, the data suggest that the apical membrane is Cl^- impermeable because of abnormal regulation of the Cl^- channel. However, the data also suggest that β -adrenergic agonists bind to their receptors, that cAMP accumulation is intact and that the K^+ channel, which is involved in Ca -dependent stimulus-response coupling, is activated normally. Thus, we speculate that the Cl^- impermeability of CF epithelia results from a defect in the metabolism of some intracellular mediator, the presence of an intracellular inhibitor of the channel, or an abnormal regulatory site on the Cl^- channel itself.

We thank Phil Karp, Timothy R. Ruppert, Robin Wicks, Theresa Butwell and Vanessa Krumbholz for technical assistance. We especially thank the patients with cystic fibrosis and their families for allowing us to obtain the tissue required for these studies. This work was supported by grants from the NHLBI (HL-29851) and NIADDK (AM-27651), the Cystic Fibrosis Foundation (National organization and Rainbow Chapter) and the American Heart Association. M.J.W. is an Established Investigator of the American Heart Association.

- Widdicombe, J. H., Welsh, M. J. & Finkbeiner, W. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6167-6171 (1985).
- Welsh, M. J. *Science* **232**, 1648-1650 (1986).
- Welsh, M. J. in *Physiology of Membrane Disorders* (eds Andreoli, T. E., Hoffman, J. F., Fanestil, D. D. & Schultz, S. G.) 751-766 (Plenum, New York, 1986).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
- Shoemaker, R. L., Reckemmer, G. & Frizzell, R. A. *Biophys. J.* **49**, 158a (1986).
- Smith, P. L., Welsh, M. J., Stoff, J. S. & Frizzell, R. A. *J. Membrane Biol.* **70**, 217-226 (1972).
- Al-Bazzaz, F. J. *Am. Rev. respir. Dis.* **123**, 295-298 (1981).
- Sato, K. & Sato, F. *J. clin. Invest.* **73**, 1763-1771 (1984).
- Welsh, M. J. & McCann, J. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8823-8826 (1985).
- Lowry, O. H., Rosebrough, N. H., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265-275 (1951).
- Liedtke, C. M. & Tandler, B. *Am. J. Physiol.* **247**, C441 (1983).

Role of the HTLV-III/LAV envelope in syncytium formation and cytopathicity

Joseph Sodroski, Wei Chun Goh, Craig Rosen,
Kathryn Campbell & William A. Haseltine*

Department of Pathology, Dana-Farber Cancer Institute, Harvard Medical School and *Department of Cancer Biology, Harvard School of Public Health, Boston, Massachusetts 02115, USA

Acquired immune deficiency syndrome (AIDS) is characterized by marked depletion of the T4^+ helper subset of T cells¹⁻³. The aetiological agent of the disease, the human T-lymphotropic virus type III (HTLV-III)/lymphadenopathy-associated virus (LAV), specifically kills T4^+ cells *in vitro*⁴⁻¹⁰. Part of this specificity for the T4^+ population resides in the relative efficiency with which HTLV-III infects these cells, as a result of a specific interaction between the T4 molecule and the virus envelope glycoprotein¹¹⁻¹³. In addition, the cytotoxic consequences of HTLV-III replication are dependent on cell type, as certain lymphoid and myeloid cells can be productively infected without notable cytopathic effect^{14,15}. Here we investigate the basis for the specific cytotoxicity of the virus, and report that high-level expression of the HTLV-III envelope gene induces syncytia and concomitant cell death in T4^+ cell lines but not in a B-lymphocyte line. Syncytium formation depends on the interaction of envelope-expressing cells with neighbouring cells bearing surface T4 molecules. These results explain, at least in part, the specific cytopathic effect of HTLV-III infections.

The investigation of the envelope gene as a determinant of HTLV-III cytopathic effect was motivated by recognition that the exterior glycoprotein of several enveloped viruses is responsible for syncytium formation and contributes to the cytotoxic effects of infection¹⁶⁻²³. Moreover, previous findings ruled out a direct cytopathic role for the proteins unique to HTLV-III, including those encoded by the *src*, 3' *orf* and *tat*-III (*trans*-activator) genes. Viruses deleted for either or both of the *src* and 3' *orf* genes replicate in and kill T4^+ cells^{24,25}. Moreover, T4^+ cell lines that express constitutively a functional *tat*-III gene product are viable²⁶.

To determine whether the HTLV-III envelope gene was involved in the induction of syncytia and cell death, plasmids were constructed that produced high levels of the envelope glycoprotein. The construction of such plasmids was made possible by the finding that expression of the envelope gene directed by the HTLV-III long terminal repeat (LTR) is dependent on two other HTLV-III genes, *tat*-III and *art*²⁶⁻²⁹. Accordingly, these two functions were supplied in the initial experiments. The presence of the *tat*-III gene was guaranteed by using cells that express constitutively the *tat*-III gene product (Jurkat-*tat*-III and Raji-*tat*-III)³⁰. The *art* function was supplied by using plasmids that were designed to express both the *art* and *env* proteins simultaneously (pIIIenv3-1 and pIIIenv3-2; ref. 29) (Fig. 1).

The T4^+ Jurkat-*tat*-III cell line³⁰ was used as a recipient in these experiments. The gp160 and gp120 HTLV-III envelope

Received 9 January; accepted 6 June 1986.

- Quinton, P. M. *Nature* **301**, 421-422 (1983).
- Knowles, M. R. *et al. Science* **221**, 1067-1070 (1983).
- Yankaskas, J. R., Knowles, M. R., Gatz, J. T. & Boucher, R. C. *Lancet* **i**, 954-956 (1985).

Table 1 Syncytium formation following transfection of cells

Transfected DNA	Added antibody	Co-cultivated cells (% T4 ⁺)	Syncytia per 10 ⁷ cells	Transfected DNA	Added antibody	Co-cultivated cells (% T4 ⁺)	Syncytia per 10 ⁷ cells
a, Jurkat-<i>tat</i>-III cells				b, Jurkat cells			
None	None	None	<1,000	ptat-I	None	None	23 ± 12
pIIEnv3-1	None	None	38,000 ± 10,000	ptat-I	None	None	2,005 ± 120
pIIEnv3-2	None	None	29,000 ± 8,000	+ pIenv3	None	None	1,952 ± 80
pIIEnv3-D1	None	None	<1,000	ptat-I	None	None	30 ± 18
pIIEnv3-FS1	None	None	<1,000	+ pIenv3-FS2	None	None	
pIIEnv3-FS2	None	None	<1,000				
pIIEnv3-FS1	None	None	76,000 ± 18,000				
+ pIIEnv3-FS2							
pIIEnv3-D1	None	None	58,000 ± 13,000	c, Raji-<i>tat</i>-III cells			
+ pIIEnv3-FS2				None	None	None	<1,000
pIIEnv3-D1	OKT8	None	65,000 ± 15,000	pIIEnv3-1	None	None	<1,000
+ pIIEnv3-FS2				pIIEnv3-D1	None	None	<1,000
pIIEnv3-D1	OKT4	None	51,000 ± 12,000	+ pIIEnv3-FS2	None	Raji (0)	<1,000
+ pIIEnv3-FS2				pIIEnv3-D1	None	Raji- <i>tat</i> -III (0)	<1,000
pIIEnv3-D1	OKT4A	None	2,000 ± 2,000	+ pIIEnv3-FS2	None	HUT78 (30)	11,000 ± 4,000
+ pIIEnv3-FS2				pIIEnv3-D1	None	C8166-45 (>95)	54,000 ± 8,000
pIIEnv3-D1	p982	None	53,000 ± 16,000	+ pIIEnv3-FS2	None	Jurkat (80)	38,000 ± 3,000
+ pIIEnv3-FS2				pIIEnv3-D1	None	Jurkat	36,000 ± 5,000
pIIEnv3-D1	12-8	None	55,000 ± 7,000	+ pIIEnv3-FS2	OKT4	Jurkat	41,000 ± 7,000
+ pIIEnv3-FS2				pIIEnv3-D1	OKT8	Jurkat	<1,000
pIIEnv3-D1	Wy	None	30,000 ± 13,000	+ pIIEnv3-FS2	OKT4A	Jurkat	19,000 ± 3,000
+ pIIEnv3-FS2				pIIEnv3-D1	None	H9 (90)	21,000 ± 6,000
pIIEnv3-D1	91-1	None	35,000 ± 8,000	+ pIIEnv3-FS2	OKT8	H9	17,000 ± 2,000
+ pIIEnv3-FS2				pIIEnv3-D1	OKT4	H9	<1,000
pIIEnv3-D1	Pe	None	44,000 ± 14,000	+ pIIEnv3-FS2	OKT4A	H9	
+ pIIEnv3-FS2							
pIIEnv3-D1	Ba	None	62,000 ± 11,000				
+ pIIEnv3-FS2							
pIIEnv3-D1	K.C.	None	57,000 ± 9,000				
+ pIIEnv3-FS2							
pIIEnv3-D1	McD.	None	50,000 ± 13,000				
+ pIIEnv3-FS2							

For these experiments, ~10⁷ cells were transfected with 10 µg of each plasmid DNA using the DEAE-dextran technique³⁸. Immediately after transfection the cells were pelleted and incubated for 30 min at room temperature in 1 ml of medium containing a 1:200 dilution of monoclonal antibody OKT8, OKT4 or OKT4A (Ortho Pharmaceuticals). The cells were then suspended in 10 ml of fresh medium containing a 1:400 dilution of monoclonal antibody and incubated at 37 °C until the syncytia were counted. In subsequent experiments, the OKT4A antibody reduced syncytium formation by >90% at dilutions of 1:800. For patient antisera containing HTLV-III neutralizing activity (p982, 12-8, 91-1, Wy and Pe) as well as control patient sera (Ba, K.C. and McD.), transfected Jurkat-*tat*-III cells were grown in medium containing a 1:10 dilution of serum and syncytia counted as described below. Neutralizing activities of the p982, 12-8 and 91-1 sera were determined to be 1:64, 1:64 and 1:32, respectively, by David Ho and Martin Hirsch using methods described previously³⁴. Neutralizing activities of the Wy and Pe antisera were determined by Thomas Matthews and Dani Bolognesi to be 1:500 and 1:150, respectively, in their assay³⁶. Immediately after transfection of Raji-*tat*-III cells with the indicated plasmids, a threefold excess of untransfected cells of the type indicated was added to the culture. Values in parentheses represent the percentage of the added cells that were T4⁺, as estimated by indirect immunofluorescence¹¹. Syncytia for the Jurkat-*tat*-III and Raji-*tat*-III experiments were counted as follows. At 72-96 h post-transfection, three random fields (×80 magnification) were photographed. Approximately 10⁴ cells were counted from the photographic prints to determine the number of syncytia. The values reported represent the mean of three independent experiments, with the range of values indicated. For the experiment using Jurkat cells, 10⁷ cells were aliquoted for counting at 72 h post-transfection. The values represent the mean (and range) of three independent experiments. The syncytia observed following transfection of the ptat-I and ptat-I + pIenv3-FS1 DNAs contained less than five nuclei per syncytium; most of the syncytia observed after transfection with the ptat-I + pIenv3-FS2 DNAs contained over 20 nuclei per syncytium.

gene products³¹ were produced in these cells on transfection with either the pIIEnv3-1 or -2 plasmids (Fig. 2a, lanes 3, 4). The expression of these proteins was eliminated by either a frameshift mutation or a deletion in the region of the envelope gene that encodes the exterior glycoprotein (pIIEnv3-FS1 and -D1, Fig. 2a, lane 5; b, lane 4).

Marked syncytium formation was observed by 48 h after transfection of Jurkat-*tat*-III cultures with plasmids that express the envelope glycoprotein (Fig. 3b, c and Table 1a). In these experiments, at least 30-80-fold increases in syncytia were observed relative to control or mock-transfected cultures, representing 0.5-1% of transfected cells, a value consistent with the transient transfection frequency of human lymphoid cell lines (J.S. and

W.A.H., unpublished observations). Greater than 90% of the syncytia are unable to exclude vital dyes by 2 weeks after transfection, indicating the cytotoxic nature of the events accompanying syncytium formation (see, for example, Fig. 3h). The syncytium formation and cytopathic effect observed in experiments with the *env*-expression plasmids were indistinguishable from those seen following infection of Jurkat-*tat*-III cells with HTLV-III (Fig. 3a).

Neither syncytium formation nor cell death was noted in cultures transfected with the plasmids containing mutations in *env* that prevent expression of the gp160 and gp120 proteins (Fig. 3d, Table 1a). However, both of these plasmids (pIIEnv3-FS1 and -D1) continued to express *art* gene function, as judged

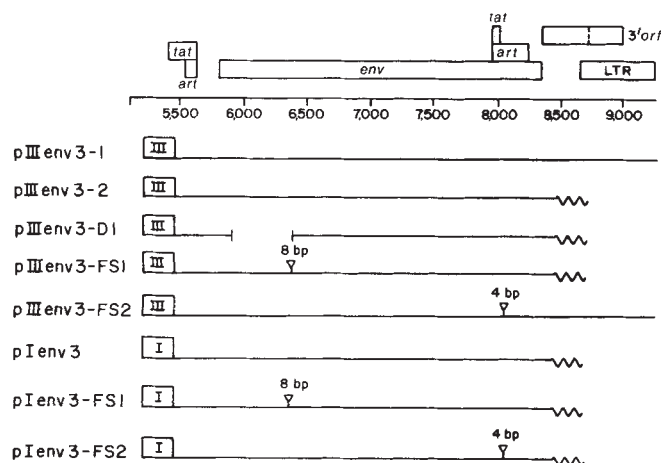


Fig. 1 Structure of the plasmids expressing *art* and *env* genes. The structure of the 3' half of the HTLV-III genome, based on the sequence of Ratner *et al.*³², is shown. The positions of the *env* gene, 3' *orf* gene and LTR are shown, together with the two coding exons of the *tat*-III and *art* genes in their respective reading frames. The position of a stop codon in the 3' *orf* gene of the parental proviral clone used in these studies, pHXBc2 (ref. 39), is denoted by a vertical broken line. In all the plasmids shown, LTR sequences from -167 to +80 for the HTLV-III LTR (boxes labelled III) or -353 to +326 of the HTLV-I LTR (boxes labelled I)⁴⁰ are located at nucleotide 5,496 of the sequence of Ratner *et al.*³². The zig-zag lines represent signals for polyadenylation and splicing derived from the simian virus 40 early region⁴¹. The eight-nucleotide insertion (8 bp) in plasmids pIIIenv3-FS1 and pIenv3-FS1 was introduced by digestion of either pIIIenv3-2 or pIenv3 with the restriction endonuclease *Stu*I and insertion of 8-bp *Xba*I linkers before transfection of *Escherichia coli*. The construction of the other plasmids has been described previously²⁹. All plasmids were banded in CsCl before transfection.

by the ability of the plasmids to complement *art*-defective proviruses for *gag* or *env* gene expression²⁹.

Of special interest is the effect of another mutation present on plasmid pIIIenv3-FS2. The frameshift mutation of this plasmid interrupts the second coding exon of the *art* gene²⁹ and results in the replacement of the 104 carboxy-terminal envelope amino acids with 57 different amino acids³². The inability of the pIIIenv3-FS2 plasmid to synthesize *env* gene products can be complemented by co-transfection with plasmids expressing the *art* gene (pIIIenv3-D1 or -FS1) (Fig. 2a, lane 6; b, lanes 5-7). An altered *env* gene product capable of inducing syncytium formation and cell death was produced in this experiment (Table 1a, Fig. 3e,f), indicating that the 104 carboxy-terminal amino acids of the HTLV-III envelope are not required for these effects.

To determine whether expression of the HTLV-III envelope in the absence of *tat*-III or *art* functions is sufficient for syncytium formation and cytopathicity, plasmids containing the HTLV-I LTR 5' to the HTLV-III *env* gene were constructed (pIenv3, pIenv3-FS1 and pIenv3-FS2; Fig. 1). One of these plasmids, pIenv3-FS2, cannot encode a functional *art* product because of a frameshift mutation in the *art* gene coding sequences. pIenv3-FS1 contains a frameshift mutation in the 5' portion of the *env* gene that encodes the exterior glycoprotein. These plasmids were transfected into Jurkat T-lymphocytes together with a plasmid (pIat-1) that expresses the *trans*-activator protein of HTLV-I³³. The level of HTLV-III envelope expression and syncytium formation observed following transfection with plasmids pIenv3 and pIenv3-FS2 was lower than that observed using the *trans*-activated HTLV-III LTR as a promoter, but was at least 60-fold higher than that observed after transfection with pIenv3-FS1 (Table 1b). The syncytia observed in this experiment were smaller (30-50 nuclei per syncytium) than those observed when the HTLV-III LTR was used as promoter (>200 nuclei per syncytium), but the viability of the syncytia was not appreciably different in the two experiments. We conclude that the HTLV-III *tat*-III and *art* genes are not required for either syncytium formation or cytopathicity consequent to *env* gene expression in T4⁺ lymphocytes.

To determine whether host cell factors might influence susceptibility to syncytium formation, a B-lymphocyte cell line that expresses constitutively the *tat*-III gene product, Raji-*tat*-III³⁰, was transfected with plasmids expressing the HTLV-III *art* and *env* genes. Raji-*tat*-III cells transfected with either pIIIenv3-1 or a combination of pIIIenv3-FS2 and pIIIenv3-D1 produced the gp160/120 envelope proteins at levels similar to those observed in Jurkat-*tat*-III cells (Fig. 2). Nonetheless, no syncytia or cytopathic effects were observed in the transfected Raji-*tat*-III cells (Table 1c).

To test whether the HTLV-III envelope glycoproteins synthesized in the Raji-*tat*-III cells were capable of inducing syncytia, the transfected Raji-*tat*-III cells were co-cultivated with various untransfected human lymphoid cell lines (Table 1c). Syncytia were observed within 72 h following addition of the human T-lymphocyte lines H9, Jurkat, C8166-45 and, to a lesser extent, HUT78 to the transfected Raji-*tat*-III cells. The syncytia formed in this experiment failed to exclude vital dyes within 7 days of their formation (data not shown). No syncytia were observed following addition of untransfected Raji or Raji-*tat*-III cells to the transfected cells. Thus, the transfected Raji-*tat*-III cells synthesized immunoprecipitable HTLV-III envelope proteins capable of inducing syncytia in a susceptible cell type.

Studies of animal retroviruses suggest that syncytium formation involves an interaction of virus-expressing cells with adjacent cells bearing receptors for the virus^{11,17-22}. The cells capable of forming syncytia following co-cultivation with Raji-*tat*-III

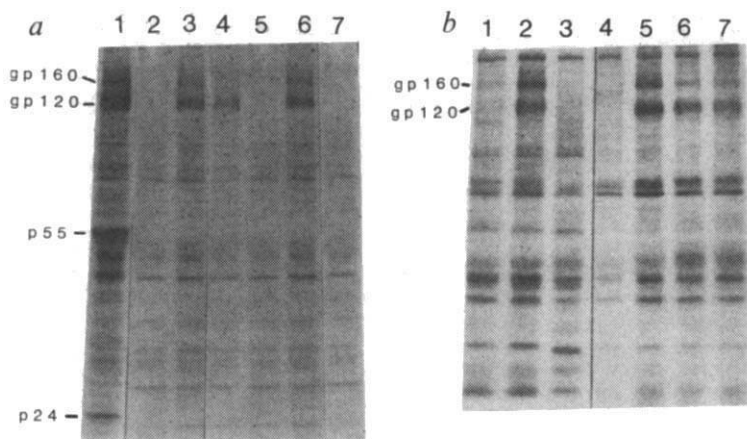


Fig. 2 Immunoprecipitation of transfected Jurkat-*tat*-III and Raji-*tat*-III cells. Jurkat-*tat*-III (a, lanes 1-7; b, lanes 4-7) or Raji-*tat*-III (b, lanes 1-3) cells were transfected by the DEAE-dextran procedure³⁷ using 10 µg of plasmid DNA; 48 h after transfection, cells were labelled with ³⁵S-cysteine and cell lysates immunoprecipitated using an AIDS patient serum, RV119. The positions of the gp160 and gp120 *env* proteins and the p55 and p24 *gag* products are noted. Transfected plasmids were pHXBc2 (a, lane 1), pBR322 (a, lane 2), pIIIenv3-1 (a, lane 3), pIIIenv3-2 (lane 4), pIIIenv3-FS1 (lane 5), pIIIenv3-FS1 plus pIIIenv3-FS2 (lane 6), pIIIenv3-FS2 (a, lane 7 and b, lane 1), pIIIenv-D1 (b, lanes 3, 4), pIIIenv-D1 plus pIIIenv3-FS2 (b, lanes 2, 5-7). Transfected Jurkat-*tat*-III cells were incubated with monoclonal antibody OKT4 or OKT4A (b, lanes 6 and 7, respectively) as described in Table 1.

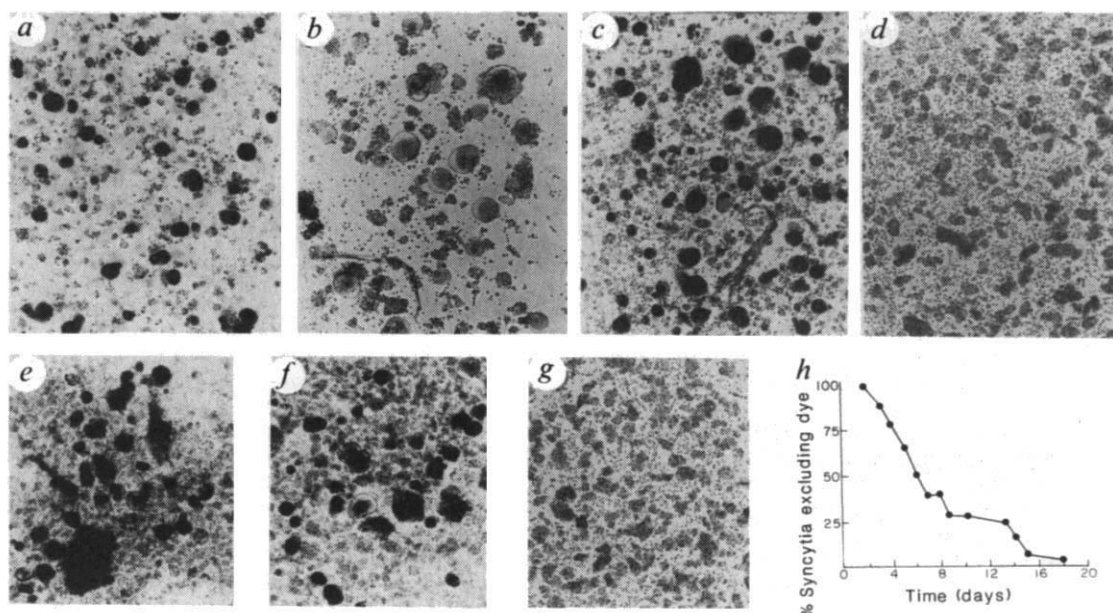


Fig. 3 Transfection of Jurkat-*tat*-III cells with plasmid DNAs. Approximately 10^7 Jurkat-*tat*-III cells were transfected with $10 \mu\text{g}$ of plasmid DNA using the DEAE-dextran procedure³⁸, and the cells were photographed at various times after transfection (a-g, $\times 53$). a, Jurkat-*tat*-III cells 5 days after transfection with plasmid pHXBc2, which generates replicating HTLV-III that kills the culture. b, c, Jurkat-*tat*-III cells 3 and 7 days, respectively, after transfection with plasmid pIIIenv3-1. The Jurkat-*tat*-III cells in d were transfected with plasmid pIIIenv3-FS1 7 days previously. e-g, Jurkat-*tat*-III cells 10 days after transfection with a combination of pIIIenv3-D1 and pIIIenv3-FS2, followed by treatment with OKT8, OKT4 and OKT4A, respectively. Treatment with monoclonal antibodies was as described in Table 1. h, Percentage of Jurkat-*tat*-III cells forming syncytia that exclude Trypan blue at various times after transfection with pIIIenv3-FS1 plus pIIIenv3-FS2. The transfected culture was fed every 3 days and aliquots removed for counting. Values were obtained by examining more than 200 random syncytia for uptake of Trypan blue. The curve shown is a typical experiment, with a maximum variation of 3 days in the time required for 50% dye exclusion in three separate experiments.

transfectants expressed the T4 molecule, which has been shown to be a component of the HTLV-III receptor¹¹⁻¹³. To examine the role of T-lymphocyte markers in syncytium induction by HTLV-III, Jurkat-*tat*-III cells were transfected simultaneously with two plasmids, pIIIenv3-D1 and pIIIenv3-FS2, the combination of which leads to high-level production of the *env* protein. The cells were then incubated with monoclonal antibodies directed against T8, T4 or T4A. Only the monoclonal antibody against T4A inhibited syncytium formation in the transfected Jurkat-*tat*-III cells (Fig. 3e-g, Table 1a). The absence of syncytia in the presence of the monoclonal antibody cannot be attributed to inhibition of *env* gene synthesis as levels of immunoprecipitable gp160/120 envelope proteins were similar regardless of the monoclonal antibody used for incubation (Fig. 2b, lanes 6, 7). The formation of syncytia following co-cultivation of envelope-expressing Raji-*tat*-III cells with T4⁺ lymphocytes (Jurkat and H9) could also be inhibited by monoclonal antibodies against T4A but not by antibodies against T8 or T4 (Table 1c). Previous studies demonstrated that monoclonal antibodies against T4A but not against T4 or T8 could inhibit the interaction of the HTLV-III envelope with the T4 molecule¹³. These results suggest that the T4 epitope that interacts with the HTLV-III envelope is a necessary component in syncytium formation.

The ability of antisera from normal individuals and from HTLV-III-infected patients to inhibit syncytium formation by the HTLV-III envelope gene was assayed. The sera tested included some that previously were found to contain relatively high antibody titres that inactivate HTLV-III in virus neutralization tests³⁴⁻³⁶. All the patient sera tested contain antibodies to envelope glycoproteins as judged by radioimmune precipitation tests (not shown). None of the patient sera significantly inhibited syncytium formation by the HTLV-III envelope gene compared with control sera (Table 1).

The results presented here demonstrate that expression of the HTLV-III envelope results in marked syncytium formation and concomitant cell death indistinguishable from the effects of infection with HTLV-III itself. Syncytium formation does not occur in a cell line lacking the T4 molecule nor does it occur in the presence of antibodies known to interfere with the interaction between the HTLV-III envelope and T4. These results suggest that syncytium formation resulting from an interaction of the HTLV-III envelope with the T4 molecule is the basis for at least part of the specific cytotoxicity of HTLV-III for T4⁺ cells. It is also likely that cell death will result from the expression of the HTLV-III envelope protein in an isolated cell that exhibits a high density of T4 surface antigen via an autofusion reaction that disrupts normal membrane function. The marked cytopathic effect of HTLV-III relative to most other retroviruses might be the consequence of several factors: the greater avidity of binding of the exterior glycoprotein for the receptor, efficient fusion activity of HTLV-III envelope domains, and high-level expression of virion components consequent to *trans*-activation.

Cells expressing the HTLV-III envelope glycoprotein, regardless of the level of expression of the T4 molecule, can serve to nucleate the formation of syncytia with uninfected cells that bear the T4 surface protein. This effect may contribute to the marked depletion of T4⁺ lymphocytes observed in AIDS patients despite the small percentage of peripheral blood lymphocytes actively producing viral messages and protein at any given time³⁷.

The dependence of the cytopathic effect of HTLV-III on the interaction of the envelope glycoprotein with a T4 surface protein provides an explanation for the ability of the virus to infect but not kill cells that do not express high levels of the T4 molecule. Dalglish *et al.*¹¹ observed that higher titres of anti-T4 antibodies are required to block infection by vesicular stomatitis

virus pseudotypes of HTLV-III than are necessary to inhibit syncytium formation, and suggested that more T4 molecules are required for syncytium formation than for infection by HTLV-III.

The antisera from HTLV-III-infected patients that we have tested exhibit at best slight inhibition of *env* gene-mediated syncytium formation or cytopathicity, even though they react with the envelope glycoprotein and are reported to inactivate virion infectivity³⁴⁻³⁶. Weiss *et al.*³⁵ also observed that most patient sera failed to inhibit syncytium formation. Inhibition of syncytium formation by antibodies is reported to be critical for the protective response to infection by other enveloped viruses²³. Failure to inhibit the fusion activity of the virus may account for the progression of HTLV-III-related diseases in the presence of antibodies reactive against the envelope glycoproteins, as cell-to-cell transmission of the virus may not be blocked in most patients.

Definition of both the binding and fusion domains of the HTLV-III envelope glycoprotein will be important for the development of strategies for protection against HTLV-III infection and consequent immune cell depletion.

We thank Drs George Shaw, Beatrice Hahn, Flossie Wong-Staal and Robert Gallo for the gift of the pHXBc2 provirus; Mary Fran McLane, Myron Essex, Dani Bolognesi, David Ho and Martin Hirsch for gifts of antisera; and Tanya Parks for manuscript preparation. W.A.H. is supported by a State of Massachusetts contract and by NIH grant CA21082. J.S. is a Special Fellow of the Leukaemia Society of America and is also the recipient of NIH grants CA40658 and CA42597. C.R. is the recipient of NIH postdoctoral fellowship CA07580 and NIH Contract FOD-0757. Both J.S. and C.R. are recipients of grants from the American Foundation for AIDS Research.

Received 14 April; accepted 4 June 1986.

- Seligman, M. *et al.* *New Engl. J. Med.* **311**, 1286-1292 (1984).
- Friedman-Kein, A. E. *et al.* *Ann. intern. Med.* **96**, 693-700 (1982).
- Gottlieb, M. S. *et al.* *New Engl. J. Med.* **305**, 1425-1431 (1981).
- Gallo, R. C. *et al.* *Science* **224**, 497-500 (1984).
- Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
- Sarngadharan, M. G., Popovic, M., Bruch, L., Shupbach, J. & Gallo, R. C. *Science* **224**, 506-508 (1984).
- Barre-Sinoussi, F. *et al.* *Science* **220**, 868-871 (1983).
- Schupbach, J. *et al.* *Science* **224**, 503-505 (1984).
- Klatzmann, D. *et al.* *Science* **225**, 59-63 (1984).
- Vilmer, E. *et al.* *Lancet* **i**, 753-756 (1984).
- Dalglish, A. *et al.* *Nature* **312**, 763-767 (1984).
- Klatzmann, D. *et al.* *Nature* **312**, 767-768 (1984).
- McDougal, J. S. *et al.* *Science* **231**, 382-385 (1986).
- DeRosi, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 4297-4301 (1986).
- Gartner, S. *et al.* *Science* **223**, 215-219 (1986).
- Poste, G. *Int. Rev. Cytol.* **13**, 157-252 (1972).
- Chatterjee, S. & Hunter, E. *Virology* **107**, 100-108 (1970).
- Diglio, C. A. & Ferrer, J. F. *Cancer Res.* **36**, 1056-1067 (1976).
- Klement, V., Rowe, W. P., Hartley, J. W. & Pugh, W. E. *Proc. natn. Acad. Sci. U.S.A.* **63**, 753-758 (1969).
- Nagy, K., Clapham, P., Cheingsong-Popov, R. & Weiss, R. A. *Int. J. Cancer* **32**, 321-328 (1983).
- Redmond, S., Peters, G. & Dickson, C. *Virology* **133**, 393-402 (1984).
- Pinter, A. *et al.* *J. Virol.* **57**, 1048-1054 (1986).
- Richardson, C. D. & Choppin, P. W. *Virology* **131**, 518-532 (1983).
- Sodroski, J. *et al.* *Science* **231**, 1549-1553 (1986).
- Terwilliger, E., Sodroski, J., Rosen, C. & Haseltine, W. *J. Virol.* (in the press).
- Dayton, A., Sodroski, J., Rosen, C., Goh, W. C. & Haseltine, W. *Cell* **44**, 941-947 (1986).
- Arya, S., Guo, C., Josephs, S. F. & Wong-Staal, F. *Science* **229**, 69-74 (1985).
- Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. A. *Science* **229**, 74-77 (1985).
- Sodroski, J. *et al.* *Nature* **321**, 412-417 (1986).
- Rosen, C. A., Sodroski, J. G., Campbell, K. & Haseltine, W. A. *J. Virol.* **57**, 379-384 (1986).
- Allan, J. S. *et al.* *Science* **228**, 1091-1093 (1985).
- Ratner, L. *et al.* *Nature* **313**, 277-284 (1985).
- Sodroski, J., Rosen, C., Goh, W. C. & Haseltine, W. *Science* **228**, 1430-1434 (1985).
- Ho, D., Rota, T. & Hirsch, M. *New Engl. J. Med.* **312**, 649-650 (1985).
- Weiss, R. A. *et al.* *Nature* **316**, 69-72 (1985).
- Robert-Guroff, M., Brown, M. J., Gallo, R. C. *Nature* **316**, 72-74 (1985).
- Harper, M. E., Marselle, L. M., Gallo, R. C. & Wong-Staal, F. *Proc. natn. Acad. Sci. U.S.A.* **83**, 772-776 (1986).
- Queen, C. & Baltimore, D. *Cell* **33**, 741-748 (1983).
- Fisher, A. C., Collalti, E., Ratner, L., Gallo, R. C. & Wong-Staal, F. *Nature* **316**, 262-265 (1985).
- Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3618-3622 (1983).
- Mulligan, R. C., Howard, B. H. & Berg, P. *Nature* **277**, 108-114 (1979).

A chromosomal rearrangement in a *P. falciparum* histidine-rich protein gene is associated with the knobless phenotype

Laura G. Pologé & Jeffrey V. Ravetch

DeWitt Wallace Research Laboratory, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

The significant morbidity and mortality associated with *Plasmodium falciparum* malaria results, in part, from the sequestration of parasitized erythrocytes in postcapillary venules, which may protect the parasite from splenic clearance^{1,2} and contribute to the pathogenesis of cerebral malaria³. This sequestration has been linked to the expression of parasite-induced knob structures on the surface of the infected erythrocyte which mediate the cytoadherence phenomenon^{4,5}. While knobs are necessary for cytoadherence, they are not sufficient, requiring both parasite- and host-encoded proteins⁶⁻¹⁰. Spontaneous mutants of *P. falciparum* have been isolated from *in vitro* cultures which lack the ability to express knobs and fail to cytoadhere¹¹. A histidine-rich protein has been described which is associated with the knobby phenotype¹² and may be a constituent of the knob¹³. We now report the isolation of complementary DNA clones for a knob-associated histidine-rich protein (KAHRP) and demonstrate that in knobless mutants the gene for this protein has undergone a rearrangement, resulting in a deletion in the 3' coding sequence. Moreover, the chromosome to which the KAHRP gene maps is rearranged in these mutants, producing a telomeric location of the truncated gene. These observations explain the loss of expression of the messenger RNA and protein in such mutants and may explain the loss of the knob itself. The implications for the generation of spontaneous mutations in the parasite by this novel mechanism are discussed.

A cDNA library constructed to *P. falciparum* strain FcR-3 at the trophozoite stage was screened with a 1.8-kilobase (kb) *EcoRI* fragment derived from the histidine-rich protein gene of *Plasmodium lophurae*¹⁴ under reduced stringency (25% formamide, 10% dextran sulphate, 5×SSC, 7 mM Tris pH 7.6, 1×Denhardt's buffer, 25 µg ml⁻¹ salmon sperm DNA at 40 °C; final wash = 0.1×SSC, 0.1% SDS, 40 °C). cDNA clones which cross-hybridized with the avian gene were characterized further by restriction enzyme and Northern blot analysis to knobby (K⁺) and knobless (K⁻) RNA. Three overlapping cDNA clones (Fig. 1a) were characterized by DNA sequence analysis and shown to encode an open reading frame containing multiple polyhistidine sequences varying between 6 and 9 contiguous histidine residues (data not shown), a primary structure analogous to the *P. lophurae* gene¹⁴. These same clones were identified in the library using a cDNA probe to unique sequences in K⁺ mRNA constructed by subtractive hybridization¹⁵ between K⁺ and K⁻ clonal isolates. Northern blot analysis (Fig. 1b, c) demonstrated that stable mRNA transcripts of 4.2 kb accumulate in K⁺ but not K⁻ isolates and are maximally expressed in trophozoites. Antibodies generated to the recombinant protein expressed in *Escherichia coli* recognize a protein (in FCR-3) of 80,000-90,000 relative molecular mass in K⁺ but not K⁻ isolates (Pologé and Ravetch, in preparation). This expression pattern is in agreement with the expression of a knob-associated histidine-rich protein described previously^{12,16}. Independent clonal derivatives of K⁺ parasites from a wide range of geographical isolates demonstrated expression of this RNA only in K⁺ isolates (see Fig. 1b). Thus, by virtue of its histidine-rich amino-acid sequence and expression in K⁺ and not K⁻ isolates, these clones represent cDNA isolates for a KAHRP gene.

To investigate how the accumulation of stable transcripts for this gene has been lost in K⁻ isolates, DNA was isolated from K⁺ and K⁻ isolates, analysed with multiple restriction endonu-

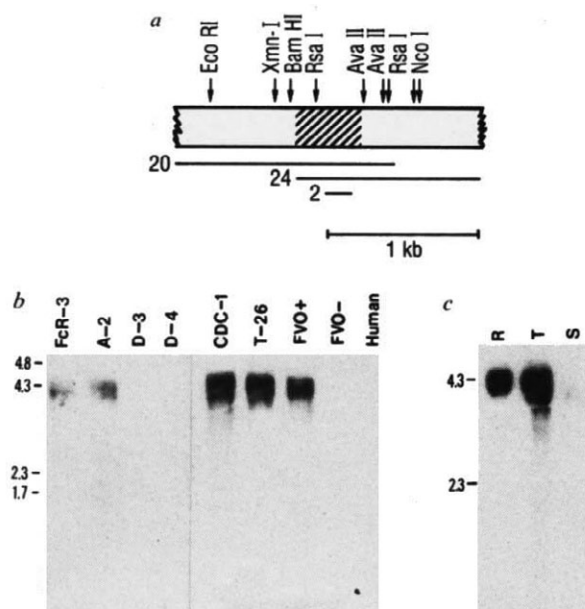


Fig. 1 *a*, Restriction map and cDNA clones for a knob-associated histidine-rich protein of *P. falciparum*. Cross-hatching, polyhistidine encoding sequences. Three overlapping cDNA clones are indicated with their numerical designation. *b*, Northern blot analysis of RNA isolated from multiple knobby (K^+) and knobless (K^-) isolates. Size markers are derived from *P. falciparum* and human rRNA sizes. FcR-3 is the non-clonal Gambian line established by Trager²⁵ and A-2 (K^+), D-3 (K^-) and D-4 (K^-) are clonal derivatives of that line²²; FVO⁺ is a K^+ line cloned from a Vietnam isolate (ref. 26) and FVO⁻ a K^- clone derived from it²⁷; CDC-1 (ref. 28) is a Honduran isolate (K^+) and T-26 is a Tanzanian isolate (K^+). Two RNA species appear to be resolved for this KAHRP gene migrating at 4.1 and 4.2 kb. *c*, Stage specificity of KAHRP gene expression. R, rings; T, trophozoites; S, schizonts. The faint hybridization seen in the schizont lane may represent contamination of schizonts with trophozoites during synchronization.

Methods. *a*, A cDNA library constructed in the *Pst*I site of pUC-9 (ref. 29) was screened and the resulting clones characterized as described in the text. Overlapping clones were identified by restriction enzyme analysis and the fragment sizes verified by co-migration with genomic DNA. DNA sequence analysis was performed using the chemical degradation method of Maxam and Gilbert³⁰. *b*, *P. falciparum* was grown in synchronous culture as described elsewhere²⁶ to obtain populations enriched for trophozoites (*b*) or rings and schizonts (*c*). 1 μ g of total RNA was fractionated on agarose-formaldehyde gels, transferred to nitrocellulose and hybridized with nick-translated cDNA probes (specific activity 2×10^8 c.p.m. per μ g), under stringent conditions (50% formamide, 10% dextran sulphate, $5 \times$ SSC, 7 mM Tris pH 7.6, $1 \times$ Denhardt's, 25μ g ml⁻¹ salmon sperm DNA; final wash = $0.1 \times$ SSC, 0.1% SDS, 52 °C). These conditions were used for all stringent hybridization studies described. The 1.4-kb *Pst*I insert of clone 20 was used in *b* and the 250-base-pair (bp) *Pst*I insert of clone 2 for *c*.

cleases and probed with the cDNA clones illustrated in Fig. 1a. As shown in Fig. 2, the restriction fragments encoding this cDNA sequence are conserved in K^+ parasites from two geographical isolates, as well as in clonal derivatives of these isolates. A single *Hind*III fragment of 10.5 kb is detected by these cDNA clones in K^+ strains (lanes 1, 3), while two *Xmn* fragments of 8.1 and 3.4 kb are detected when clone 20 is used as a probe, as a result of an internal *Xmn* site in this cDNA clone (lanes 5, 6, 9). Similar studies on K^+ isolates from Honduras (CDC-1) and Tanzania (T-26) reveal that the gene for this KAHRP is conserved (not shown). However, a DNA rearrangement has occurred in K^- isolates, resulting in more rapidly migrating species. The *Hind*III fragment is 6.1 kb in D-4 and 7.2 kb in FVO⁻ (Fig.

2, lanes 2, 4). Similarly, the 5' *Xmn* fragment in FVO⁻ is unrearranged, while the 3' fragment, encoding the polyhistidine sequences, now migrates as a diffuse band of 2.3 ± 0.25 kb (Fig. 2, lane 10). K^- isolates D-3 and D-4 have lost the 5' *Xmn* site, resulting in a single rearranged *Xmn* fragment of 5.2 kb (Fig. 2, lanes 7, 8). This rearrangement has resulted in a deletion of DNA sequences corresponding to the 3' coding sequence of this gene. Restriction enzyme mapping of the strains shown in Fig. 2 with *Eco*RI, *Bam*HI, *Xmn*I, *Ava*II and *Hind*III, singly and in combination, using probes derived from both the 5' and 3' sequences of the KAHRP gene, demonstrated that the breakpoint of the deletion in different K^- isolates varies by several hundred nucleotides. For clones D-3 and D-4, derived from FcR-3, the deletion breakpoint results in the loss of all histidine-encoding sequences, while in FVO⁻, derived from a Vietnam isolate, the breakpoint retains polyhistidine sequences (see Fig. 3c). The rearranged DNA fragment observed in these K^- isolates migrates as a diffuse band (Fig. 2, lanes 2, 4, 7, 8, 10), implying that the DNA fragment that is generated varies in length in the different K^- isolates. In addition, our restriction mapping studies suggest that a clustering of restriction enzyme cleavage sites has been introduced 3' of this gene in K^- isolates (data not shown).

The finding that a common restriction site had been introduced 3' of the rearranged gene in K^- isolates and the heterogeneity of the DNA fragment in these mutants suggested that the gene for the KAHRP in K^- isolates might have become relocated to the end of a chromosome. To test this hypothesis directly, DNA from FVO⁺ and FVO⁻ parasites was subjected to *Bal*31 digestion for increasing lengths of time, then digested with the restriction enzyme *Hind*III. Southern blot analysis with the cDNA probe revealed that the 10.5-kb *Hind*III fragment is unaffected by *Bal*31 digestion in the K^+ isolate (Fig. 3a), while the 7.2-kb *Hind*III fragment in the K^- isolate migrates as a 4.3-kb fragment after 30 min of *Bal*31 digestion (Fig. 3b). In other experiments, the 10.5-kb fragment remains undigested in this K^+ isolate at 60 min while the DNA from the K^- isolate shows no detectable hybridization at 35 min (not shown). Similar analysis of A-2, D-3 and D-4 revealed increased susceptibility to *Bal*31 in the K^- isolates (data not shown). From the size of the truncated gene fragment in the K^- isolates, we can estimate that the rearrangement which has occurred in these mutants has relocated this gene within 2.5 kb of a telomere. Figure 3c summarizes findings regarding the structure of the KAHRP gene in K^+ and K^- isolates.

These observations suggested that there may be a chromosomal rearrangement in the K^- isolates involving the KAHRP gene. *P. falciparum* chromosomes were separated by pulse-field gradient electrophoresis¹⁷ and hybridized with the cDNA clones isolated for the KAHRP (Fig. 4). The KAHRP gene maps to chromosome 2, which appears to be invariant in the K^+ isolates used. However, the K^- isolates demonstrate a more rapidly migrating chromosome 2 which hybridizes with reduced intensity to the cDNA probe for this gene, due to the deletion within the KAHRP gene. As has been noted previously^{18,19}, *P. falciparum* isolates differ in their karyotype with respect to the size and number of chromosomes. For the clonal isolates FVO⁺ and FVO⁻, and also for the clonal isolates A-2 (K^+), D-3 (K^-) and D-4 (K^-), the only karyotypic difference observed is in chromosome 2. FcR-3, the non-clonal parent for A-2, D-3 and D-4, is perplexing in its karyotype when compared with its clonal isolates for chromosomes 1 and 4. The only chromosomal rearrangement, however, observed consistently for all K^+ and K^- isolates, is seen for chromosome 2. Hybridization with the *P. falciparum* ribosomal RNA gene probe²⁰ detected no difference in the clonal pairs (not shown), nor was there a difference when the gene for the glycophorin-binding protein GBP130 (ref. 21) was used. These results demonstrate that the DNA rearrangement found in K^- isolates for the KAHRP gene is reflected at the chromosomal level, and may represent a large deletion from chromosome 2.

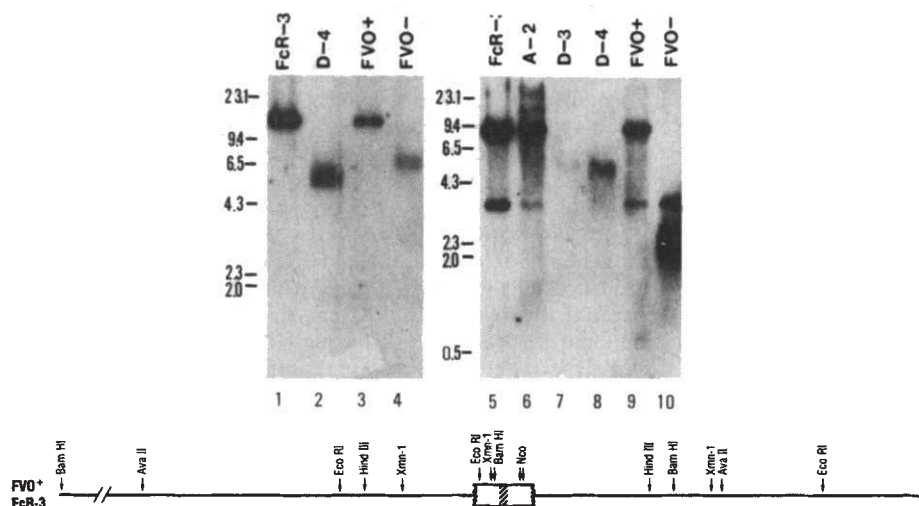
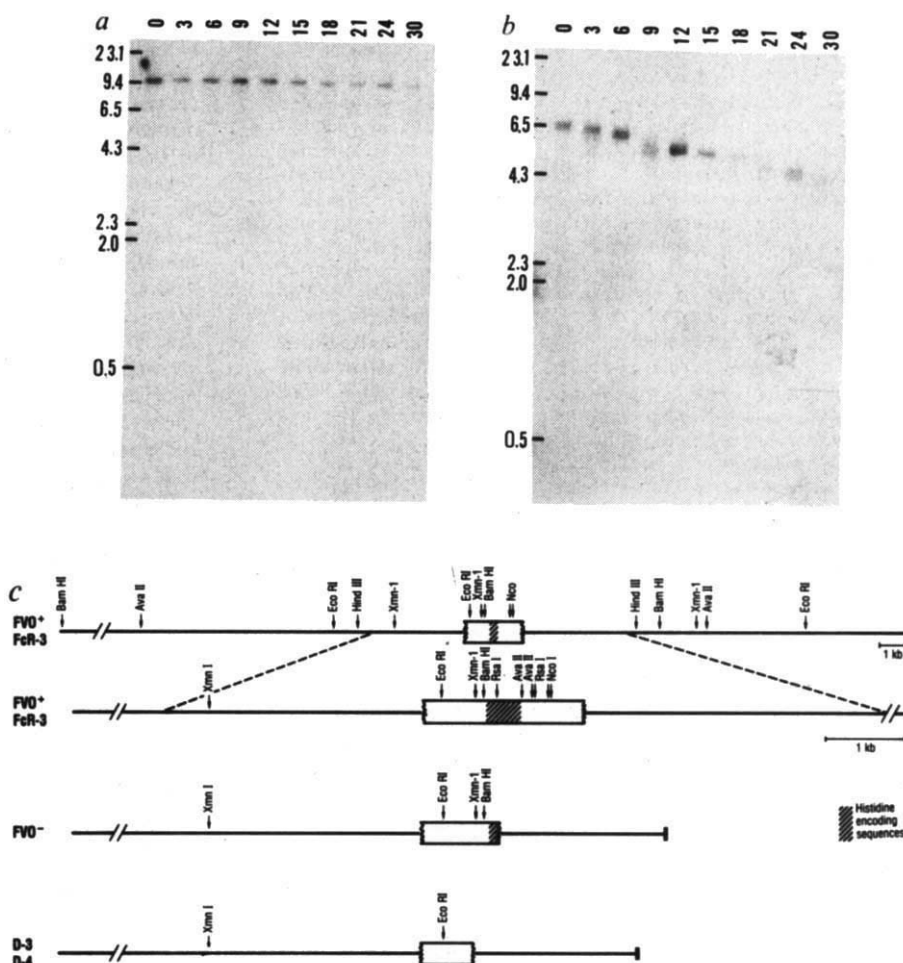


Fig. 3 *a, b*, *Bal31* digestion of the KAHRP gene in K^+ and K^- isolates. Numbers at the top of each lane refer to time in minutes. *c*, Summary of restriction maps of the KAHRP in K^+ and K^- isolates. The breakpoints in FVO $^-$, D-3 and D-4 are approximate, mapping to within the *Bam*HI-*Ava*II fragment for FVO $^-$ and the *Xmn*I-*Eco*RI fragment for D-3 and D-4. The solid bar at the end of the maps for the K^- isolates indicates the end of a chromosome, as determined by the *Bal31* experiments described above. Hatching indicates polyhistidine-encoding sequences.

Methods. *a, b*, FVO $^+$ (*a*) and FVO $^-$ (*b*) DNA was obtained from these isolates as described elsewhere³², and digested with *Bal31* according to the manufacturer's (NEB) specifications in a reaction containing 5 μ g DNA, 1 U *Bal31* in a total reaction volume of 175 μ l. At the indicated time points, aliquots were brought to 20 mM EGTA, heated to 70 $^{\circ}$ C for 5 min and ethanol-precipitated. The DNA samples were digested with *Hind*III, fractionated on 1% agarose, Tris-acetate EDTA gels and transferred as described previously³¹. Hybridization under stringent conditions was performed using a 250-bp *Pst*-*Eco*RI fragment derived from clone 20 (see Fig. 1*a*). The rate of precession of the *Bal31* was determined to be 80 bp min $^{-1}$, based on digestion of λ DNA cut with *Hind*III under the conditions used for these experiments. The increased sensitivity of the gene in the K^- isolate is seen. Analogous experiments performed with A-2 (K^+), D-3 (K^-) and D-4 (K^-) revealed a similar pattern of increased *Bal31* sensitivity (not shown) in the K^- isolates. *c*, The restriction maps were generated using the enzymes indicated both singly and in combination. Sites within the cloned sequences were verified by DNA sequencing analysis.

These results demonstrate that the loss of expression of a KAHRP gene in knobless mutants is associated with a chromosomal rearrangement and deletion of part of the gene itself. The molecular mechanism which could account for the telomeric location for the truncated gene is under investigation. We do not know the extent of the deletion on chromosome 2, nor have we determined whether a reciprocal translocation involving telomeric exchange of DNA from chromosome 2 to some other

Fig. 2 Southern blot analysis³¹ of the KAHRP gene in DNA isolated from K^+ and K^- isolates. Lanes 1-4, *Hind*III digestions of DNA isolated from FcR-3 (lane 1), D-4 (lane 2), FVO $^+$ (lane 3) and FVO $^-$ (lane 4) probed with a 5' *Pst*I-*Eco*RI fragment of 250 bp derived from clone 20; lanes 5-10, *Xmn*I digests of FcR-3 (lane 5), A-2 (lane 6), D-3 (lane 7), D-4 (lane 8), FVO $^+$ (lane 9) and FVO $^-$ (lane 10) probed with the *Pst*I insert of clone 20. Under the stringent conditions of hybridization and washing used, only one hybridizing fragment is detected for *Hind*III and two fragments for *Xmn*I, as expected for K^+ DNA. The genomic restriction map for this KAHRP gene is presented below the autoradiogram of the Southern blot. The cloned and sequenced region of the gene is indicated by the open box; hatching indicates the histidine-encoding sequences.



chromosome has occurred. The K^- phenotype occurs spontaneously in culture²², resulting in parasites which are otherwise wild type in their growth. If these mutants occur *in vivo*, they would presumably be selected against by splenic clearance, as has been shown²³. It is possible that K^- mutants accumulate in culture because of the absence of the negative selection exerted by the spleen *in vivo*. Thus, this chromosomal rearrangement may reflect a more general mechanism for mutation by the

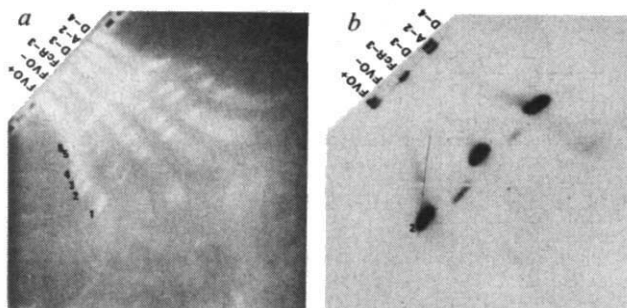


Fig. 4 Pulse-field gradient electrophoresis of *P. falciparum* chromosomes isolated from K^+ and K^- isolates. *a*, The ethidium bromide-stained gel pattern for various K^+ and K^- isolates; *b*, an autoradiogram of the gel in *a* hybridized with a cDNA clone to the KAHRP gene.

Methods. Agarose blocks containing *P. falciparum*-infected erythrocytes were prepared as described previously^{17,18} using parasites grown in culture to parasitaemias of 10–20%. Chromosome separation was performed using the method of Schwartz and Cantor¹⁷ in 1% agarose, 0.5×TBE gels electrophoresed at 210 V for 22 h with a pulse frequency of 75 s in north-south and east-west directions. Following electrophoresis and staining in 1 µg ml⁻¹ of ethidium bromide, the gels were soaked in 0.25 M HCl for 1 h before transfer to nitrocellulose as described by Southern³¹. Hybridization under stringent conditions was performed using the nick-translated *Pst*I fragment of clone 20. The strains are described in Fig. 1 legend. The numbering of the chromosomes is according to van der Ploeg *et al.*¹⁸, with chromosome 1 migrating at 700 kb and chromosome 6 at 2,000 kb.

parasite, resulting in the diverse karyotypes found in nature which could be of selective advantage in parasite survival. Although mechanistically this rearrangement may resemble the relocation of certain variable surface glycoprotein genes to telomeres in trypanosomes²⁴, the result of the K^+ to K^- rearrangement is the loss of expression of a KAHRP gene. Analysis of these K^+ to K^- mutations will provide an opportunity to determine the generality of chromosomal rearrangement in plasmodia and may yield insights into its mechanism.

This work was supported in part by grants to J.V.R. from the Rockefeller Foundation, the United Nations Development Programme-World Bank World Health Organization Special Programme for Research and Training in Tropical Diseases and US Army Contract DAMD17-85-C-5177. J.V.R. is a Pew Scholar and a recipient of an award from the Pew Scholars Program in the biomedical sciences, supported by the Pew Memorial Trust. L.G.P. is a fellow of the Schepp Foundation and supported in part by an Exxon Fellowship Training Grant in Molecular Biology. We thank Lex van der Ploeg for assistance with our PFG studies, Margaret Perkins, Lex van der Ploeg and Peter Model for critical reading of this manuscript, M. Samuels, H. Valsamis and A. Pavlovic for technical assistance, and Karen Yates and Mary Wentzler for secretarial assistance.

Received 28 January; accepted 23 April 1986.

1. Miller, L. H. *Am. J. trop. Med. Hyg.* **18**, 860–865 (1969).
2. Barnwell, J. W., Howard, R. J. & Miller, L. H. *Ciba Fdn Symp.* **94**, 117–136 (1983).
3. Edington, G. M. *Ann. trop. Med. Parasit.* **48**, 300–306 (1954).
4. Trager, W., Rudzinska, M. A. & Bradbury, P. C. *Bull. Wild Hlth Org.* **35**, 883–885 (1966).
5. Luse, S. A. & Miller, L. H. *Am. J. trop. Med. Hyg.* **20**, 655–660 (1971).
6. Udeinya, I. J., Schmidt, J. A., Aikawa, M., Miller, L. H. & Green, I. *Science* **213**, 555–557 (1981).
7. Udeinya, I. J., Graves, P. M., Carter, R., Aikawa, M. & Miller, L. H. *Exp. Parasit.* **56**, 207–214 (1983).
8. Roberts, D. D. *et al. Nature* **318**, 64–66 (1985).
9. Leech, J. H., Barnwell, J. W., Miller, L. H. & Howard, R. J. *J. exp. Med.* **159**, 1567–1575 (1984).
10. Aley, S. B., Sherwood, J. A. & Howard, R. J. *J. exp. Med.* **160**, 1585–1590 (1984).
11. Schmidt, J. A. *et al. J. clin. Invest.* **70**, 379–386 (1982).
12. Kilejian, A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4650–4653 (1979).
13. Leech, J. H., Barnwell, J. W., Aikawa, M., Miller, L. H. & Howard, R. J. *J. Cell Biol.* **98**, 1256–1264 (1984).
14. Ravetch, J. V., Feder, R., Pavlovic, A. & Blobel, G. *Nature* **312**, 616–620 (1984).
15. Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. *Nature* **308**, 149–153 (1984).
16. Vernot-Hernandez, J.-P. & Heidrich, H.-G. *Molec. Biochem. Parasit.* **12**, 337–350 (1984).

17. Schwartz, D. C. & Cantor, C. R. *Cell* **37**, 67–75 (1984).
18. van der Ploeg, L. H. T. *et al. Science* **229**, 658–660 (1985).
19. Kemp, D. J. *et al. Nature* **315**, 347–350 (1985).
20. Langsley, G., Hyde, J. E., Goman, M. & Scaife, J. G. *Nucleic Acids Res.* **11**, 8703–8717 (1983).
21. Ravetch, J. V., Kochan, J. & Perkins, M. *Science* **227**, 1593–1597 (1985).
22. Trager, W. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 6527–6530 (1981).
23. Langreth, S. G. & Peterson, E. *Infect. Immunity* **47**, 760–766 (1985).
24. van der Ploeg, L. H. T., Schwartz, D. C., Cantor, C. R. & Borst, P. *Cell* **37**, 77–84 (1984).
25. Jensen, J. B. & Trager, W. *Am. J. trop. Med. Hyg.* **27**, 743–746 (1978).
26. Trager, W. & Jensen, J. B. *Science* **193**, 673–675 (1976).
27. Gritzmacher, C. A. & Reese, R. T. *Science* **226**, 65–67 (1984).
28. Bhasim, V. K. & Trager, W. *Am. J. trop. Med. Hyg.* **33**, 534–537 (1984).
29. Kochan, J., Perkins, M. & Ravetch, J. V. *Cell* **44**, 689–696 (1986).
30. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
31. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
32. Goman, M. *et al. Molec. Biochem. Parasit.* **5**, 391–400 (1982).

Long-range restriction site mapping of mammalian genomic DNA

William R. A. Brown & Adrian P. Bird

MRC Mammalian Genome Unit, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK

Molecular analysis of many problems in genetics would be facilitated by the ability to construct restriction site maps of long stretches of genomic DNA and to directly place genes on these maps. Pulsed-field gradient gel electrophoresis allows measurement of the size of DNA fragments up to at least 2,000 kilobase pairs (kb) long^{1,2} and we have used this technique here to map sites for one class of infrequently cutting restriction enzyme over a total of 1,500 kb of mouse genomic DNA. The sites for these enzymes tend to be clustered in the genome. These clusters may correspond to the short stretches of C+G-rich unmethylated DNA³ often associated with mammalian genes.

To construct long-range maps of genomic DNA, enzymes are needed which cut the DNA infrequently. Sites for enzymes which have a 6-base-pair (bp) recognition sequence composed entirely of inter-strand C-G base pairs and which include one or more CpG dinucleotides (C-G enzymes) are rare in the mammalian genome because the DNA is low in G+C and because the CpG dinucleotide occurs at only 20–25% of the frequency expected from base composition⁴. In addition, the CpG dinucleotide is often methylated at cytosine and this inhibits cleavage by many C-G enzymes (see Table 1). It occurred to us that cleavable sites for methylation-sensitive C-G enzymes should not be distributed randomly with respect to one another in mammalian DNA but should be concentrated in short stretches of genomic DNA called HTF islands³; this is predicted because HTF islands are 65% C+G, show no CpG suppression, and are unmethylated (see Table 1). As the rare inter-island sites are likely to be methylated and therefore resistant to cleavage, our calculations suggest that almost all cleavable sites for methylation-sensitive C-G enzymes will be in HTF islands. Two lines of evidence which support this prediction are shown in Fig. 1. When DNA of high relative molecular mass (M_r) is restricted with *Nae*I, *Sac*II or *Sma*I and analysed by pulsed-field gradient gel electrophoresis, the DNA in each digest is distributed from ~50 kb up to the limit of resolution, which in this experiment was at 650 kb; about 30% of the DNA in the sample was above this limit. The *Sac*II recognition sequence (Table 1) includes two CpG dinucleotides while the recognition sequences for *Nae*I and *Sma*I contain only one, yet all three enzymes gave a similar distribution of fragments. This result suggests that these enzymes are cutting in regions of the genome which show no CpG suppression. The observation that the double and single digests all contained a similar distribution of fragments suggests that the sites for the different enzymes are close to one another.

The standard strategy for mapping of restriction sites in genomic DNA involves a series of single and double enzyme

digests, gel electrophoresis, blotting and hybridization. In general, however, this technique does not work where the sites for different enzymes are clustered. An alternative mapping strategy is to regard a particular HTF island as a landmark and to use pulsed-field gel electrophoresis to measure the distance from sites in that known island to the flanking sites located tens or hundreds of kilobases away. We used this strategy to map about three HTF islands in the mouse genome. Two of these islands were isolated at random from genomic DNA³ and one was from the 5' end of the dihydrofolate reductase gene (*dhfr*)^{5,6}. Although our results were obtained using mouse DNA, this approach can be applied to the DNA of other mammals.

The first region of the genome studied was that which flanks island HTF 9. This island was isolated from mouse DNA³ and is known to be associated with transcripts that occur in several mouse tissues (P. Lavia, unpublished results). Island HTF 9 has two sites each for the enzymes *NarI*, *SacII* and *SmaI* and three sites for *NaeI* (Fig. 2a). We digested high-*M_r* DNA with each of the four enzymes. To check that the digest was complete at the sites within the island, we restricted an aliquot of each digest with *EcoRI* and, after conventional agarose gel electrophoresis, analysed the products with a probe which encompasses island HTF 9 (Fig. 2b). As no trace of the full-sized *EcoRI* fragment remained in any of the double digests, we concluded that each enzyme could cleave completely within island HTF 9. This check on digestion is necessary for the construction of an unambiguous map because *NaeI*, *NarI* and *SacII* show site-dependent heterogeneity in their reaction rates (see ref. 7 and below). We then fractionated DNA that had been digested with the infrequently cutting enzymes by pulsed-field gradient gel electrophoresis, and analysed the fractionated DNA by filter hybridization. The probes in this experiment flank island HTF 9, thus the size of any hybridizing restriction fragment localizes the restriction site with respect to the HTF island. The gels were run at a shorter pulse time than those shown in Fig. 1 to allow better resolution of the fragments of lower *M_r*. Analysis with the left-hand probe (Fig. 2c) revealed two hybridizing fragments in the *NaeI* digest: a small weakly hybridizing fragment and a larger fragment close to the unfractionated DNA. The latter fragment was more accurately sized on a gel run at a longer pulse time, and measured ~375 kb (not shown). In addition, there was a *NarI* fragment of 140 kb, two *SacII* fragments (30 and 140 kb) and a weakly hybridizing *SmaI* fragment (140 kb). On the right-hand side of the island (Fig. 2d) we observed a *NaeI* fragment of ~105 kb and *NarI*, *SacII* and *SmaI* fragments each ~40 kb long. Together these results yield the restriction site map around island HTF 9 shown in Fig. 2e. This map also indicates the estimated error in the restriction fragment sizes which we have measured from our data. A limitation of this sort of map is that it may be incomplete. Because we are making measurements about a single point, we are unable to map distal to sites which are cut to completion. A striking feature of the analysis shown in Fig. 2c is the presence of more than one hybridizing fragment in the *NaeI* and *SacII* digests, which means that although island HTF 9 is completely cleaved by these enzymes, it is flanked on one side by *NaeI* and *SacII* sites which are only partially cleaved. These sites are clustered with sites for other tested enzymes, thus they would be expected to represent uncharacterized HTF islands. A possible explanation is that the site-dependent kinetics of cleavage by *NaeI* and *SacII*⁷ is responsible for the partial cleavage. The possibility cannot be excluded, however, that incomplete cleavage is due to partial methylation at these sites. Whatever the cause, partial cleavage increases the amount of data from the experiments.

We have used the same strategy to construct long-range maps around island HTF 12 (Fig. 3) and the HTF island at the 5' end of *dhfr* (Fig. 4). Two technical points were important in the construction of the long-range maps (Figs 3e, 4e). First, island HTF 12 lacks any *NarI* sites and so we were unable to locate the ends of the 375-kb *NarI* fragment including island HTF 12.

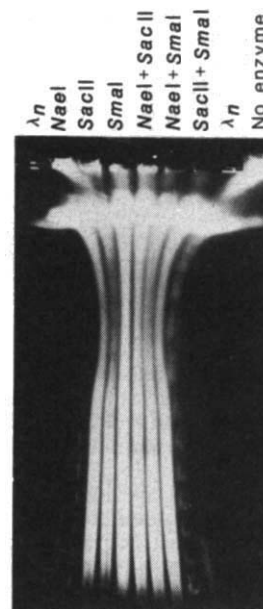


Fig. 1 Pulsed-field gradient gel electrophoresis of high-*M_r* mouse DNA restricted with either *NaeI*, *SacII* or *SmaI* alone or in combination. The *M_r* markers are multimers of phage λ -c1857 DNA produced by annealing the cohesive ends in free solution (λ_n).

Methods. For the preparation of high-*M_r* DNA, BALB/c mouse thymocytes were encapsulated in agarose beads at a concentration of 5×10^7 per ml of beads¹¹. The beads, washed free of paraffin, were washed once in phosphate-buffered saline (PBS), once in Tris-EDTA buffer (TE: 10 mM Tris-HCl, 1 mM EDTA pH 7.5), then the nuclei were lysed with detergent by resuspending them at a concentration of 2×10^7 cells ml⁻¹ in 10 mM Tris, 0.05 M EDTA, 1% SDS pH 8.0 at room temperature. The suspension became like a jelly, presumably because of the release of DNA from cells at the periphery of individual beads, and so was thoroughly mixed by forcing it through the 2-mm orifice of a pipette. When the suspension of beads had become less viscous, the beads were washed twice in TE buffer and once in an EDTA/lauroyl sarcosine buffer (NDS: 0.5 M EDTA, 10 mM Tris-HCl, 1% lauroyl sarcosine pH 9.5) and finally resuspended in 2 vol. NDS. Proteinase K (Boehringer) was added to a final concentration of 500 μ g ml⁻¹ and the suspension of beads was incubated at 55 °C for 48 h. This digest was then made to 50% (v/v) with glycerol and stored at -20 °C. For restriction enzyme digestion of DNA, beads were washed in TE buffer (pH 7.8) and residual proteinase K inactivated with 5 mM phenylmethylsulphonyl fluoride. The beads were restricted as follows: 120 μ l of beads restricted with 15 U of *NaeI* in a total volume of 150 μ l; 240 μ l of beads restricted with 62 U *SacII* in a total volume of 300 μ l; and 360 μ l of beads restricted with 70 U *SmaI* in a total volume of 450 μ l. These samples were incubated at the recommended temperature for 5 h, then washed in TE. The slurry of washed beads containing *SacII*-digested DNA was divided into two equally sized aliquots, one of which was digested with 14 U *NaeI*. The slurry of beads containing *SmaI*-digested DNA was subdivided into three equally sized aliquots, one of which was further restricted with 14 U *NaeI* and another with 50 U *SacII*. The second round of digests continued for 5 h and the beads were then washed and loaded onto the electrophoresis slots. An aliquot of washed beads (120 μ l) was incubated for 10 h in restriction buffer at 37 °C as a control for nonspecific nuclease degradation. These beads were washed and loaded into the right-most well. The pulsed-field gradient gel electrophoresis chamber was based on the design of Carle and Olson² but was modified to take a 20 $\times$ 20 cm gel plate. Construction details may be obtained from W.R.A.B. The DNA sample to be fractionated was loaded into a well of dimensions 8 $\times$ 2 $\times$ 6.5 mm. The total depth of the agarose gel was 7.5 mm. The agarose (Sigma type II; Medium EEO) concentration was 1.5% (w/v) and the electrophoresis buffer was 0.5 $\times$ TAE (TAE is 0.04 M Tris-acetate, 0.001 M EDTA). The run time in this experiment was 26 h, buffer temperature 8 °C, pulse time 30 s.

Second, we were unable to completely cleave the *NarI* site in the *dhfr* HTF island and so could not place the *NarI* sites which flank this gene. The consistent failure to cleave this particular *NarI* site may be the result of either site-dependent heterogeneity in the reaction rate of this enzyme or partial methylation at this site. The former explanation is strongly supported by our finding that the same *NarI* site in a genomic clone of *dhfr* is not cleaved completely even in the presence of a 40-fold excess of enzyme (data not shown).

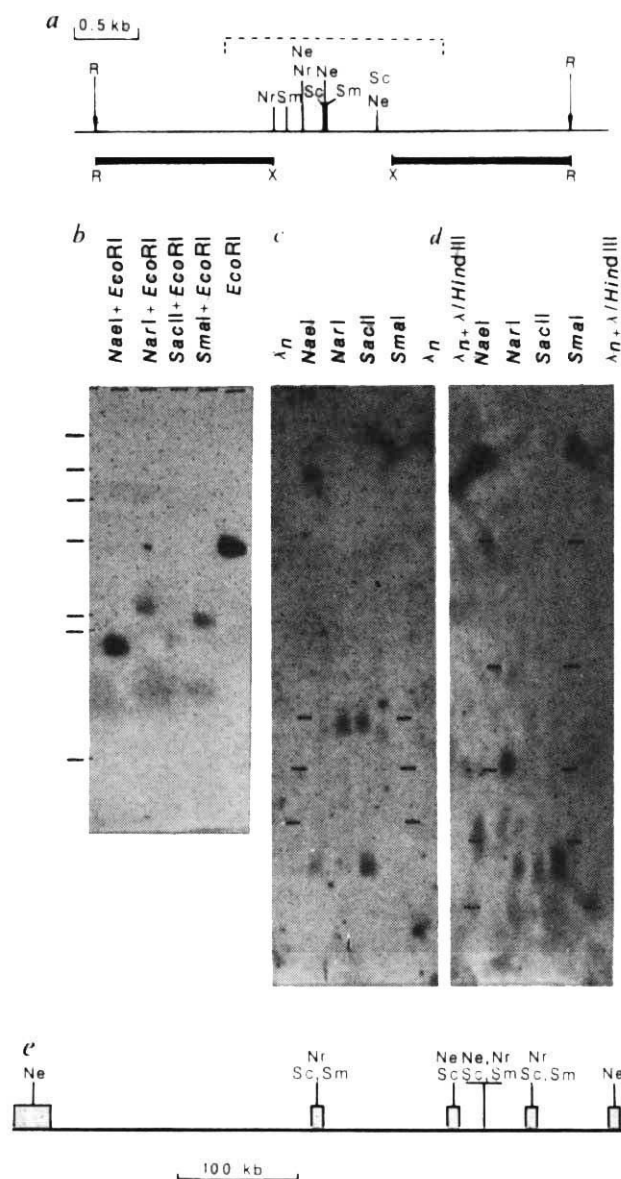


Fig. 2 Long-range mapping around island HTF 9. **a**, Restriction site map of island HTF 9. Ne, *NaeI*; Nr, *NarI*; Sc, *SacII*; Sm, *SmaI*; X, *XhoI*; R, *EcoRI*. The probes used in the long-range mapping are indicated by the solid bars beneath the map. The broken line above the map indicates the extent of unmethylated DNA. **b**, Analysis of *NaeI*, *NarI*, *SacII* and *SmaI* digestion at island HTF 9. High-*M*, DNA that had either been restricted with *NaeI*, *NarI*, *SacII* or *SmaI* or incubated in restriction buffer without enzyme as described in Fig. 1 legend, was washed and restricted with *EcoRI*. Digests were fractionated by conventional 1% agarose gel electrophoresis and analysed using a 2.6-kb *HindIII* fragment including the whole island as a probe. The standards used in this gel were products of phage λ DNA restricted with *HindIII*. Some fragments are weaker than expected or undetectable on the blot because of their small size and also because low-*M*, DNA fragments diffuse out of the heads during restriction and washing. The sizes of the restriction fragments seen in the digests were estimated as *EcoRI* 3.8 kb, *NaeI*+*EcoRI* 1.8 kb, *NarI*+*EcoRI* 2.5 kb, *SacII*+*EcoRI* 1.8 kb and *SmaI*+*EcoRI* 2.2 kb. These values are consistent with each of the infrequently cutting enzymes cleaving at one or more of the respective sites in this island. **c**, Mapping to the left of island HTF 9. High-*M*, DNA restricted with *NaeI*, *NarI*, *SacII* or *SmaI* was fractionated by pulsed-field gradient gel electrophoresis, transferred to GeneScreen Plus and analysed by hybridization using the left-hand *XhoI*-*EcoRI* fragment shown in **a**. Conditions of electrophoresis: pulse time 22.5 s, run time 20 h, buffer temperature 18°C, 300 V. **d**, Mapping to the right of island HTF 9. High-*M*, DNA restricted and fractionated as described above was transferred to nitrocellulose and analysed by hybridization using the right-hand *XhoI*-*EcoRI* fragment shown in **a**. Conditions of electrophoresis: pulse time 8.4 s, run time 27 h, buffer temperature 8°C, 300 V. The bars marked on the autoradiographs in **c** and **d** indicate the positions of the oligomers of phage λ c11857 DNA run in the flanking tracks of these gels. The broken bars in **d** show the positions of the 23-kb fragment of a λ DNA/*HindIII* digest also run in the flanking tracks of this gel. The positions of these markers were determined by projecting the images of the ethidium bromide-stained gels onto the autoradiographs. **e**, Long-range map around island HTF 9. The map was constructed from the data shown in **c** and **d** and other similar experiments. In particular, there is a *NaeI* site ~375 kb to the left of island HTF 9; this corresponds to a band towards the top of the *NaeI* track in the autoradiograph shown in **c**. This band is beyond the markers on this particular gel and is not accurately sized at a short pulse time, therefore the size of this *NaeI* fragment was estimated from two other similar experiments in which the gel was run with a pulse time of 30 s. The error associated with fragment sizes is caused by the diffuseness of the hybridization signals. This uncertainty is indicated by the square brackets at the relevant restriction sites. The cluster of sites corresponding to HTF island 9 is underlined. **Methods.** High-*M*, DNA was restricted with *NaeI*, *NarI*, *SacII* or *SmaI* and analysed as described in Fig. 1 legend. The conditions of the pulsed-field gradient gel electrophoresis were chosen to optimize the resolution in the *M*, range corresponding to the restriction site clusters. Filter transfer¹² of DNA from the pulsed-field gels was as follows: gels were incubated with gentle shaking in 0.25 M HCl for 40 min, rinsed in distilled water and incubated with gentle shaking in 0.5 M NaOH, 1.5 M NaCl for 60 min with one change. The gel was finally neutralized in 0.5 M Tris-HCl, 3 M NaCl pH 5.0 for 90 min with two changes. Transfer was done in 10×SSC. Filters were hybridized to 20 ng of probe radiolabelled¹³ to 2×10^5 c.p.m. μg^{-1} in a volume of 15 ml. GeneScreen Plus was hybridized according to conditions recommended by the manufacturer; nitrocellulose was hybridized in 5× Denhardt's 5×SET, 0.1% SDS, 0.1% sodium pyrophosphate, 10% dextran sulphate at 65°C. Filters were washed down to 0.2×SSC, 0.1% SDS, 0.1% sodium pyrophosphate at 65°C and autoradiographed for 2 weeks at -70°C using an intensifying screen.

Table 1 Distribution of potential C-G endonuclease sites between bulk DNA and the HTF island fraction in mouse DNA

Enzyme	Site	No. of CpGs	m ⁵ C sensitivity*	No. of potential sites per genome ($\times 10^4$)		Per cent sites in islands	Sites per island
				Bulk DNA	Islands		
<i>SmaI</i>	CCCGGG	1	+	4.8	3.5	42	1.2
<i>NaeI</i>	GCCGGC	1	+	4.8	3.5	42	1.2
<i>NarI</i>	GGCGCC	1	+	4.8	3.5	42	1.2
<i>SacII</i>	CCGCGG	2	+	1.2	3.5	74	1.2

For the calculation of number of sites per haploid genome, we took the base composition of bulk DNA as 40% G+C, the CpG deficiency as 25% of the expected level, and the haploid genome size as 3×10^9 bp. HTF islands were assumed to comprise 1% of genomic DNA, to have a base composition of 65% G+C, and to show no deficiency of CpG. HTF islands of 500 to 2,000 bp have been found in mouse DNA, so for the purpose of calculation an average length of 1,000 bp has been assumed; this implies that there are 3×10^4 HTF islands per haploid genome and that they occur on average every 100 kb. The values for the number of sites in bulk DNA refer to the number of potential sites and do not take into account the effect of cytosine methylation which reduces the number of sites that can be cleaved by the enzyme.

* *SmaI* is a C-G enzyme which was known to be inhibited by methylation but nothing was known about the effects of methylation on the cleavage by other C-G enzymes. We tested *NaeI*, *NarI* and *SacII* by attempting to cleave the ribosomal DNA of *Xenopus laevis* either isolated in a molecular clone or in the sperm genome. Sequence data⁹ indicated that there are many sites for each of these enzymes within the untranscribed spacer of the ribosomal DNA and these could be cleaved in the molecular clone. The rDNA in sperm is 97% methylated at CpG (ref. 10) and was largely refractory to cleavage (not shown), thus we concluded that these enzymes are inhibited greatly by cytosine methylation.

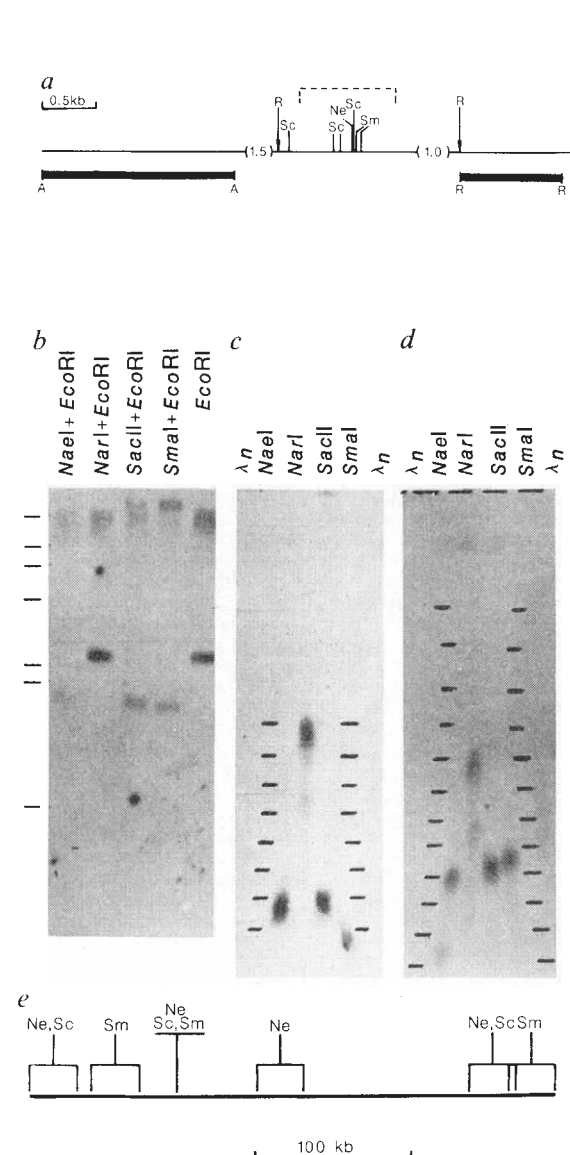


Fig. 3 Long-range mapping around island HTF 12. **a**, Restriction site map of island HTF 12. The probes used in the analysis shown in **c** and **d** are indicated by the bars beneath the map. The left-hand probe is an *AccI* fragment (abbreviated A). The right-hand probe is an *EcoRI* fragment. The broken lines indicate the extent of unmethylated genomic DNA. **b**, High- M_r DNA either restricted with *NaeI*, *NarI*, *SacII* or *SmaI* or incubated in restriction buffer without enzyme, was washed and *EcoRI*-restricted. Digests were analysed after conventional agarose gel electrophoresis using the central *EcoRI* fragment shown in **a** as a probe. Markers were the products of λ DNA restricted with *HindIII*. The sizes of the restriction fragments seen in the digests were estimated as (kb): *EcoRI* 2.45, *NaeI*+*EcoRI* 1.76, *NarI*+*EcoRI* 2.53, *SacII*+*EcoRI* 1.7, *SmaI*+*EcoRI* 1.7. These values are consistent with *NaeI*, *SacII* and *SmaI* cleaving once within the island. **c**, Mapping to the left of island HTF 12. High- M_r DNA restricted with either *NaeI*, *NarI*, *SacII* or *SmaI* was fractionated by pulsed-field gradient gel electrophoresis and analysed by filter hybridization using the left-hand *AccI* fragment shown in **a**. Electrophoresis conditions: pulse time 30 s, run time 23 h, buffer temperature 15 °C, 300 V. **d**, Mapping to the right of island HTF 12. High- M_r DNA restricted and fractionated as above was analysed by filter hybridization to the right-hand *EcoRI* fragment. The electrophoresis conditions were as described for **c**. The bars on the autoradiographs show the positions of the oligomers of phage λ c1857 DNA run in the flanking tracks of these gels. **e**, Long-range map around island HTF 12. Construction details are given in Fig. 2 legend. The cluster of sites corresponding to island HTF 12 is underlined. Methods were as described in Fig. 2 legend except that all DNA samples were from within a single set of digestions and all filters were GeneScreen Plus.

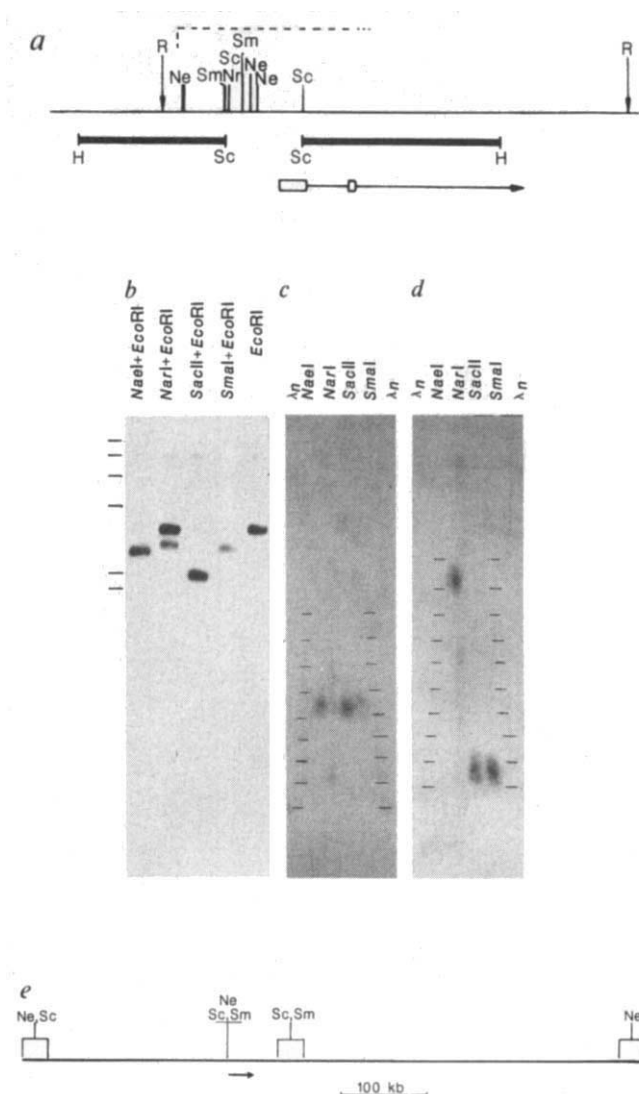


Fig. 4 Long-range mapping around the mouse dihydrofolate reductase gene (*dhfr*). **a**, Restriction site map of the HTF island at the 5' end of *dhfr*. The positions of the probes used in the analysis illustrated in **b-d** are indicated by the bars beneath the map. Also shown is the 5' end of the primary transcript. The first two exons are boxed. The broken lines above the map indicate the extent of unmethylated genomic DNA. H, *HindIII*. **b**, High- M_r DNA either restricted with *NaeI*, *NarI*, *SacII* or *SmaI*, or incubated in restriction buffer without enzyme, was washed and *EcoRI*-restricted. Digests were analysed after conventional agarose gel electrophoresis using the right-hand *SacI*-*HindIII* fragment shown in **a** as a probe. The products of phage λ DNA restricted with *HindIII* served as markers. The fragment sizes are as follows (kb): *EcoRI* 3.33, *NaeI*+*EcoRI* 2.82, *NarI*+*EcoRI* 3.33 and 2.93, *SacII*+*EcoRI* 2.24, *SmaI*+*EcoRI* 2.82. These values indicate that *NaeI*, *SacII* and *SmaI* all cleaved within this HTF island. The presence of an intact *EcoRI* fragment in the *EcoRI*+*NarI* double digest indicates that *NarI* did not cut completely in this island; this was also seen in a second experiment. **c**, Mapping to the left of *dhfr*. High- M_r DNA restricted with either *NaeI*, *NarI*, *SacII* or *SmaI* was fractionated by pulsed-field gradient gel electrophoresis and analysed by filter hybridization using the left-hand *SacII*-*HindIII* fragment described in **a**. Electrophoresis conditions: pulse time 30 s, run time 22 h, buffer temperature 16 °C, 320 V. **d**, Mapping to the right-hand side of *dhfr*. High- M_r DNA restricted and fractionated as above was analysed by filter hybridization using the right-hand *SacII*-*HindIII* fragment illustrated in **a**. Electrophoresis conditions: pulse time 30 s, run time 23 h, buffer temperature 15 °C, 300 V. The bars on the autoradiographs show the positions of the oligomers of phage λ c1857 DNA run in the flanking tracks of these gels. **e**, Long-range map around *dhfr*. The map was constructed on the basis of the data shown in **c** and **d**. The left-hand *SmaI* site was not mapped because a possible hybridizing fragment was obscured by a scratch on the autoradiograph shown in **d**. In a second experiment the *SmaI* digest was only partially complete at the *dhfr* HTF island. The position of the 34-kb *dhfr* primary transcript is indicated by the arrow. The cluster of sites corresponding to the *dhfr* HTF island is underlined. DNA samples were all from within a single set of digestions and all filters were GeneScreen Plus. For other details see Fig. 2 legend.

In the present study we have used pulsed-field gradient gel electrophoresis to map cleavable sites for the enzymes *NaeI*, *NarI*, *SacII* and *SmaI* over ~1.5 million base pairs of the mouse genome. As predicted, the sites available for cleavage by these enzymes (in other words, non-methylated sites) occur in clusters which correspond to regions rich in non-methylated CpG. Additional recognition sites for the enzymes are present outside clusters but most of the potential sites in these regions are probably not cleaved because they are methylated. Making the assumption that sites of cleavage by more than one enzyme represent HTF islands, the average distance between islands in the mapped region is 109 kb (mean of seven values). The average distance between islands has been indirectly estimated as 100 kb (ref. 3), which agrees quite well with the present value. It is clear, though, that variation about the mean is considerable.

CpG-rich islands have now been found in association with many genes (reviewed in ref. 8), and it is likely that a high proportion of the 30,000 or so HTF islands are gene-associated. As many genes with HTF islands are already available as molecular clones, it will be possible to use the strategy of mapping from HTF islands to tackle a variety of long-range mapping problems. Moreover, the putative HTF islands identified by the map data may lead directly to new genes.

We thank Chris Tyler-Smith and Howard Cooke for comments on the manuscript, Patrizia Lavia and Donald Macleod for access to unpublished data, Jean Scott for a cosmid containing the 5' end of *dhfr* and Anne Deane for typing the manuscript.

Received 13 February; accepted 23 April 1986.

- Schwartz, D. C. & Cantor, C. R. *Cell* **37**, 67-75 (1984).
- Carle, G. F. & Olson, M. V. *Nucleic Acids Res.* **12**, 5647-5664 (1984).
- Bird, A. P., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. *Cell* **40**, 91-99 (1985).
- Swartz, M. N., Trautner, T. A. & Kornberg, A. *J. biol. Chem.* **237**, 1961-1967 (1962).
- Tykocinski, M. L. & Max, E. C. *Nucleic Acids Res.* **12**, 4385-4396 (1984).
- Crouse, G. F., Simonsen, C. C., McEwan, R. N. & Schimke, R. T. *J. biol. Chem.* **257**, 7887-7897 (1982).
- New England Biolabs Cat.* (1985-86).
- Bird, A. P. *Nature* **321**, 209-213 (1986).
- Moss, T., Bosely, P. G. & Birnstiel, M. L. *Nucleic Acids Res.* **8**, 467-485 (1980).
- Macleod, D. & Bird, A. P. *Nature* **306**, 200-203 (1983).
- Cook, P. R. *EMBO J.* **3**, 1837-1842 (1984).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).

Primary sequence of a dimeric bacterial haemoglobin from *Vitreoscilla*

S. Wakabayashi*, H. Matsubara* & D. A. Webster†

* Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka 569, Japan

† Department of Biology, Illinois Institute of Technology, Chicago, Illinois 60616, USA

Vitreoscilla, a filamentous bacterium in the Beggiatoa family, synthesizes a soluble haemprotein which has two identical subunits of relative molecular mass 15,775 and two *b* haems per molecule¹. It is synthesized in relatively large quantities when the organism, a strict aerobe, is grown under hypoxic conditions². It forms a relatively stable oxygenated form which is spectrally similar to oxymyoglobin (oxyHb) and oxyhaemoglobin (oxyHb)³. The amino acid sequence of this protein has been determined and aligned to fit the helical regions of several animal and plant globins. This alignment is consistent with its being a structural homologue of the eucaryotic haemoglobins although it diverged from the others in the N-terminal region and may lack an A-helix. It showed the maximum sequence homology (24%) with lupin leghaemoglobin (Lb). *Vitreoscilla* Hb is the first bacterial haemoglobin to be sequenced. It may function to enable the organism to survive in oxygen-limited environments by acting as an oxygen storage-trap or to facilitate oxygen diffusion.

Partially purified preparations of this *Vitreoscilla* haemprotein, which was previously called soluble cytochrome *o*, contained NADH-cytochrome *o* reductase activity, exhibited slow turnover of the NADH, oxygen consumption, and it was presumed that this protein functioned as a soluble terminal oxidase³. An oxygenated form detectable in these preparations was subsequently observed to be the predominant species of the protein present in intact cells of the bacterium during aerobic respiration⁴. The latter result and more recent experiments have provided evidence against the function of this protein as an oxidase. A true membrane cytochrome *o* was identified by photolysis of CO-liganded intact cells at -100 °C and eventually purified⁵, and a cytochrome *d* was also identified^{5,6}. The need for a soluble oxidase in the presence of these established membrane terminal oxidase became even less evident. When the CO-compound of the protein was photodissociated with a dye laser in the presence of oxygen, or when the oxygenated compound itself was photodissociated, recombination with oxygen occurred with no evidence of oxidation of the heme iron⁷. This is in contrast to results expected for a terminal oxidase and observed for cytochrome *c* oxidase⁸ but similar to what is observed for oxygen binding proteins like haemoglobin and myoglobin⁹. An infrared spectroscopic study of the stretch band for oxygen liganded to the heme iron showed that the absorption band was similar to those observed for oxymyoglobins and oxyhaemoglobins, evidence for a similar bent-end-on oxygen-type bonding¹⁰. Finally, the results reported here show that the amino acid sequence has considerable homology with that of yellow lupin Lb II. Consequently, we have reconsidered the role of this protein as a terminal oxidase and now propose that it is a haemoglobin-like protein which will accordingly be called *Vitreoscilla* Hb hereafter.

A summary of the sequence analysis of *Vitreoscilla* Hb is shown in Fig. 1. It consists of 146 amino-acid residues with a calculated relative molecular mass of 15,775 for the apoprotein, somewhat larger than the 13,000 estimated previously by SDS polyacrylamide gel electrophoresis¹. The sequence of the carboxymethylated protein was determined using trypsin and *Staphylococcus* V8 protease digestion, separation of peptides by HPLC on an ODS column, and manual and automated solid-phase Edman degradations^{11,12}. The only technical difficulty that was encountered was in deducing the sequence near the C-terminus of the polypeptide chain.

Carboxypeptidase Y released alanine, leucine, valine, glutamate, and several other amino acids from the carboxymethylated protein, but no peptide with these amino acids as C-terminus was isolated. The C-terminal tryptic peptide, T-11, which could not be recovered from the HPLC column, was obtained by precipitating the digest with dilute acetic acid and although impure was successfully sequenced, but the C-terminal amino acid could not be determined because again carboxypeptidase Y digestion released many amino acids. Chymotryptic digestion of peptide T-11 gave three peptides, and their amino-acid compositions were in agreement with the sequence of T-11. Carboxypeptidase Y digestion of peptide C-3 released a small amount of glutamic acid. Since glutamic acid does not fit the cleavage site specificity of either trypsin or chymotrypsin we assigned peptide C-3 to be the C-terminal peptide of the cytochrome. Although *Vitreoscilla* Hb is the first bacterial haemoglobin to be sequenced, a soluble *Rhizobium* haemoglobin has been reported in free-living (cultured) cells of *Rhizobium* that is unrelated to Lb and has about the same molecular size as *Vitreoscilla* Hb¹³.

The sequence of *Vitreoscilla* Hb is aligned for comparison with other haemoglobins in Table 1. It shows the greatest sequence homology with lupin Lb¹⁴. The final alignment shown was obtained from the initial computer alignment by positioning hydrophobic residues at the approximately thirty sites where internal hydrophobic residues are conserved in the normal globin structure¹⁵. This was accomplished with several

Table 1 Alignment and sequence comparison of *Vitreoscilla* Hb with several animal and plant haemoglobins

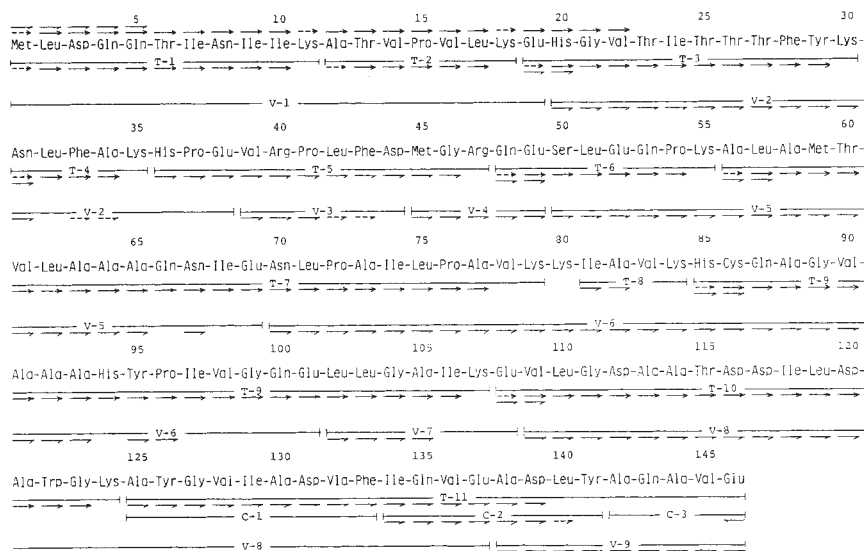
Residue no. (R)	Helix posn (H)	Haemoglobin						Residue no. (R)	Helix posn (H)	Haemoglobin					
		HHb α	HHb β	SWMb	GHb	LLb	VHb			HHb α	HHb β	SWMb	GHb	LLb	VHb
1			V	V	G	G		84					V		
2	NA1	L	H	L	L	L		85				G	T		
3	NA2	S	L	S	S	L		86	EF2	D	D	K	D	G	N
4	A1	P	T	E	A	L		87					E	V	L
5	A2	P	P	E	A	T		88	EF3			G	G	V	P
6	A3	A	E	G	A	E		89	EF4	D	N	H	K	A	A*
7	A4	D	E	F	Q	Q		90	EF5	M	L	H	M	S	I
8	A5	K	K	W	R	A		91	EF6	P	K	E	V	D	L
9	A6	T	S	Q	Q	A	M	92	EF7	N	G	A	A	A	P
10	A7	N	A	V	V	L	L*	93	EF8	N	T	E	E	T	P
11	A8	V	V	V	I	V	D	94	F1	A	F	L	M	L	V†
12	A9	K	T	L	A	K	Q	95	F2	S	A	K	K	K	K*
13	A10	A	A	H	A	S	Q	96	F3	A	T	P	A	N	K†
14	A11	A	L	V	T	S	T	97	F4	L	L	L	V	L	†
15	A12	W	W	W	W	W	I	98	F5	S	S	A	G	G	A
16	A13	G	G	A	K	E	N	99	F6	D	E	Q	V	S	V
17	A14	K	K	K	D	E	I	100	F7	L	L	S	R	V	K
18	A15	V	V	V	I	F	I	101	F8	H	H	H	H	H	H*
19	A16	G		E	A	N	K	102	F9	A	C	A	K	V	C
20	AB1	A		A	G	A	A*	103	F10	H	D	T	G	S	Q
21					N			104	FG1	K	K	K	Y	K	A
22					D			105	FG2	L	L	H	G	G	G*
23	B1	H	N	D	N	N	T	106					D	K	
24	B2	A	V	V	G	I	V	107					K	H	
25	B3	G	D	A	A	P	P*	108	FG3	R	H	K	I		
26	B4	E	E	G	G	K		109	FG4	V	V	I	P	V	V††
27	B5	Y	V	H	V	H		110	G1	D	D	P	I	K	A*
28	B6	G	G	G	G	T		111	G2	P	P	P	I	K	A
29	B7	A	G	Q	K	H	V	112	G3	V	E	N	K	E	A*
30	B8	E	E	D	D	R	L	113	G4	N	N	N	Y	Y	H*
31	B9	A	A	I	C	F	K	114	G5	F	F	L	E	F	Y†
32	B10	L	L	L	L	F	E	115	G6	K	R	L	F	P	†
33	B11	E	G	I	I	I	H	116	G7	L	L	L	F	V	I
34	B12	R	R	R	K	L	G	117	G8	L	L	L	I	V	V††
35	B13	M	L	L	H	V	V††	118	G9	S	H	G	S	K	G
36	B14	F	L	L	F	L	T†	119	G10	H	N	N	E	E	Q
37	B15	L	V	K	S	E	I	120	G11	C	V	V	A	A	E†?
38	B16	S	V	S	A	I	T	121	G12	L	L	L	I	L	L†
39	C1	F	Y	H	H	A	T	122	G13	L	V	V	I	L	L*
40	C2	P	P	P	P	P	T	123	G14	V	C	C	H	S	G
41	C3	T	W	E	E	A	F	124	G15	T	V	V	L	A	A
42	C4	T	T	T	M	A	Y†	125	G16	L	L	L	M	T	I††
43	C5	K	Q	L	A	K	K*	126	G17	A	A	H	H	E	K*
44	C6	T	R	E	A	D	N*	127	G18	A	H	S	R	R	E*
45	C7	Y	F	K	V	L	L*	128	G19	H	H	S	R	H	V*
46	CD1	F	F	F	F	F	F††	129	GH1	L	F	F	H	I	V
47	CD2	P	E	D	F	S		130	GH2	P	G	G	P	G	G*
48	CD3	H	S	R			K	131	GH3	A	K	K	G	G	D
49	CD4	F	F	F	F	F	H†	132	GH4	E	E	N	N	K	A
50	CD5		G	K	S	L	P	133	GH5	F	F	F	M	W	A
51	CD6	D	D	H	G	K	E	134	H1	T	T	G	D	S	T
52	CD7	L	L	L		G	V	135	H2	P	P	A	A	E	D
53	CD8	S	S	K			R	136	H3	A	P	D	A	E	D
54	D1		T	T		T		137	H4	V	V	A	A	L	I
55	D2		P	E		S	P	138	H5	H	Q	Q	K	N	L
56	D3		D	A	A	E	L	139	H6	A	A	G	A	S	D
57	D4		A	E	S	V	F	140	H7	S	A	A	A	A	A*
58	D5		V	M		P	D	141	H8	L	Y	M	W	W	W*
59	D6	H	M	K		Q	M	142	H9	D	Q	N	A	T	G
60	D7	G	G	A	D	N	G	143	H10	K	K	K	A	I	K
61	E1	S	N	S	D	N	R	144	H11	F	V	V	A	A	A††
62	E2	A	P	E	P	P	Q	145	H12	L	V	A	Y	Y	Y††
63	E3	Q	K	D	A	E	E*	146	H13	A	A	E	L	D	G
64	E4	V	V	L	V	L	S†	147	H14	S	G	L	A	E	V
65	E5	K	K	K	A	Q	L	148	H15	V	V	F	I	L	I†
66	E6	G	A	K	D	A	E	149	H16	S	A	R	S	A	A*
67	E7	H	H	H	L	H	Q	150	H17	T	N	K	G	I	D
68	E8	G	G	G	G	A	P	151	H18	V	A	D	A	V	V*
69	E9	K	K	V	A	G	K	152	H19	L	A	I	A	I	F†
70	E10	K	K	T	K	K	A	153	H20	T	A	A	I	K	I
71	E11	V	V	V	V	V	L†	154	H21	S	H	K	S	K	Q
72	E12	A	L	L	L	F	A	155	H22	K	K	Y	G	E	V
73	E13	D	G	T	A	K	M	156	H23	Y	Y	Y	L	M	E
74	E14	A	A	A	E	L	T	157	H24	R	H	K	E	D	A
75	E15	L	F	L	I	V	V††	158	H25			E	S	D	D*
76	E16	T	S	G	G	Y	L	159				L		A	L
77	E17	N	D	A	V	E	A	160	HC1			G			Y
78	E18	A	G	I	A	A	A††	161	HC2			Y			A
79	E19	V	L	L	V	A	A††	162	HC3			Q			Q
80	E20	A	A	K	S	I	Q	163	HC4			G			A
81	EF1	H	H	K	H	Q	N	164							V
82		V	L		L	L	E	165							E

R is an arbitrary residue number with the numbering beginning from the N-terminal of the globins with the longest N-terminal extension. H refers to the helix position in sperm whale myoglobin. The haemoglobins are: HHb α , human α -chain; HHb β , human β -chain; SWMb, sperm whale myoglobin; GHb, *Glycera* haemoglobin; LLb, lupin leghaemoglobin; VHb, *Vitreoscilla* haemoglobin. Sequence data for the first five proteins were taken directly from ref. 14, omitting three residues which were each inadvertently entered twice in the sequence for LLb presented there.

* Residues in *Vitreoscilla* Hb that match those in lupin Lb.

† Conserved, buried hydrophobic residues²³.

Fig. 1 Sequence analysis of *Vitreoscilla* Hb. T-, V- and C- refer to tryptic, *Staphylococcus* V8 protease and chymotryptic peptides, respectively. Arrows above the residues at N-terminal regions indicate the results of N-terminal sequence analysis of the apoprotein. Arrows (→), (⇌) and (←) show automated solid-phase (DITC method) Edman degradation, manual Edman degradation, and carboxypeptidase Y digestion, respectively. Dotted arrows indicate ambiguous identifications (→) or unrecovered residues by the solid-phase method (⇌).



exceptions which are probably not important (Glu at G11 and Ser at E4). The final alignment has His at F8, the proximal ligand, and the invariant Phe at CD1. The usual residue at E7, His, the distal ligand, is replaced by Gln, which has been observed in the α -chain of opossum Hb and would also be capable of forming a hydrogen bond with the bound oxygen. There are 35 matches with the 153 residues of the lupin legHb, a 24% identity per length. The number of identical residues among these various globins is given in Table 2. Note that the number of identical residues between sperm whale Mb and

Table 2 Identical residues matrix: the number of matching amino acid residues among the sequences of the globins in Table 1

	HHb α	HHb β	SWMb	GHb	LLb
HHb α					
HHb β	50				
SWMb	38	33			
GHb	30	29	32		
LLb	23	23	26	23	
VHb	16	16	14	20	34

human Hb chains is approximately the same as that between *Vitreoscilla* Hb and Lb, which is indicative of significant homology. In addition to the identical matches there are 20 similar residue matches, such as Glu/Asp, Ser/Thr, Leu/Val/Ile, Phe/Tyr, and roughly another twenty 'neutral' matches. The N-terminal region of the *Vitreoscilla* Hb does diverge from the other globins listed in Table 1, and the A-helix appears to be incomplete or lacking.

It is premature to speculate whether *Vitreoscilla* Hb shares a common evolutionary ancestor with the plant and animal haemoglobins or is the product of convergent evolution. We found no notable sequence homology between *Vitreoscilla* Hb and various cytochromes by computer analysis. Some observed similarities in the primary sequences of various globins and b-type cytochromes led Runnegar to propose that all globins are monophyletic, the common ancestor being a b-type cytochrome¹⁶. The existence of bacterial haemoglobins implies that this common ancestor would have to be a bacterial b-type cytochrome.

The following observations relate to the function of Hb in *Vitreoscilla*. The cellular haem content is dependent on the dissolved oxygen in the growth medium and increases about 50-fold when the oxygen concentration falls below about 10% atmospheric²; most or all of this is due to an increase in the *Vitreoscilla* Hb¹⁷. Species in this genus tend to live in oxygen-poor environments such as stagnant ponds and decaying veget-

able matter¹⁸, they are unable to ferment sugars, and their metabolism appears limited to aerobic oxidation of amino acids¹⁹. An adaptation to hypoxic conditions by increasing the concentration of terminal oxidases in the cellular membrane is unlikely because enzymatically active membranes, which include chloroplast, inner mitochondrial, and bacterial cell membranes, already have a protein to lipid ratio of around 3:1, in contrast to a ratio of roughly 1:1 for most animal cell membranes²⁰. An increase in membrane terminal oxidase would likely result in destabilization of the membrane. On the other hand, the concentration of haemoglobin in erythrocyte cytoplasm is more than 80% that of the crystalline protein²¹. Thus, the concentration of Hb could increase to very high concentrations in *Vitreoscilla* cytoplasm, without adversely affecting its normal metabolism, where it could then function as an oxygen storage-trap to enable the organism to survive in oxygen poor environments. Alternatively, if it were localized near the bacterial plasma membrane it could function to facilitate oxygen diffusion to the membrane terminal oxidases²². Perhaps primitive haemoglobin first appeared to play a similar role during the evolution of aerobic respiration.

The authors are indebted to Drs Russell, F. Doolittle and Peter E. Wright for invaluable assistance with the sequence homologies. Most of this work was done while D.A.W. was on a sabbatical leave in Japan supported by US NSF grant INT-8212086. Partial support also came from the US PHS NIH grant GM27085.

Received 9 January; accepted 9 May 1986

1. Tyree, B. & Webster, D. A. *J. biol. Chem.* **253**, 6988-6991 (1978).
2. Boerman, S. & Webster, D. A. *J. Gen. appl. Microbiol.* **28**, 35-43 (1982).
3. Webster, D. A. & Liu, C. Y. *J. biol. Chem.* **249**, 4257-4260 (1974).
4. Webster, D. A. & Orii, Y. *J. biol. Chem.* **252**, 1834-1836 (1977).
5. DeMaio, R. A., Webster, D. A. & Chance, B. *J. biol. Chem.* **258**, 13768-13771 (1983).
6. Webster, D. & Georgiou, C. *Fedn. Proc.* **44**, 1780 (1985).
7. Orii, Y. & Webster, D. A. *J. biol. Chem.* **261**, 3544-3547 (1986).
8. Orii, Y. *J. biol. Chem.* **259**, 7187-7190 (1984).
9. McCray, J. A. *Biochem. biophys. Res. Commun.* **47**, 187-193 (1972).
10. Choc, M. G., Webster, D. A. & Caughey, W. S. *J. biol. Chem.* **257**, 865-869 (1982).
11. Wakabayashi, S., Matsubara, H., Kim, C. H. & King, T. E. *J. biol. Chem.* **257**, 9335-9344 (1982).
12. Minami, Y., Wakabayashi, S., Imoto, S., Ohta, Y. & Matsubara, H. *J. Biochemistry (Tokyo)* **98**, 649-655 (1985).
13. Appleby, C. A. *Biochim. biophys. Acta* **172**, 88-105 (1969).
14. Lesk, A. M. & Chothia, C. *J. molec. Biol.* **136**, 225-270 (1980).
15. Perutz, M. F., Kendrew, J. C. & Watson, H. C. *J. molec. Biol.* **13**, 669-678 (1965).
16. Runnegar, B. *J. molec. Evol.* **21**, 33-41 (1984).
17. Lamba, P. & Webster, D. A. *J. Bact.* **142**, 169-173 (1980).
18. Pringsheim, E. G. *J. gen. Microbiol.* **5**, 124-149 (1951).
19. Mayfield, D. C. & Kester, A. S. *J. Bact.* **112**, 1052-1056 (1972).
20. Guidotti, G. *Ann. Rev. Biochem.* **41**, 731-752 (1972).
21. Stryer, L. in *Biochemistry*, 847 (Freeman, San Francisco, 1975).
22. Wittenberg, J. A., Bergensen, F. J., Appleby, C. A. & Turner, G. L. *J. biol. Chem.* **249**, 4057-4066 (1974).
23. Pritsyn, O. B. *J. molec. Biol.* **88**, 287-300 (1974).

Homology between IgE-binding factor and retrovirus genes

THE recent report by Toh *et al.*¹ of sequence homology between segments of a cloned gene encoding rodent IgE-binding factor (IgE-BF)² and a gene encoding Syrian hamster intracisternal A-particle (IAP)³ is an intriguing observation. However, we must express reservations about Toh *et al.*'s interpretations of the observed homology.

First, the authors claim that because the N-terminal third of the IgE-BF coding sequence shows no homology with their particular IAP gene, the IgE-BF gene must be "a hybrid gene which evolved... by integrating genes of viral origin". No evidence is presented by either Toh *et al.* or Martens *et al.*² in support of the authors' assertions that the cloned IgE-BF gene is a hybrid gene or that it even contains a portion of a non-IAP gene at all. Neither report presents any genomic blot data to indicate whether homologous genes exist in the mouse or rat genomes. Here, the conclusions of Toh *et al.* are considerably weakened by the fact that their sequence comparison crosses species lines; they really have no justification for making any statement regarding the nature of IgE-BF sequences which do not show homology to their hamster IAP sequence. Thus, their assertion that the N-terminal third of the IgE-BF amino-acid sequence is unique to IgE-BF is speculative and not supported by their data.

The IgE-BF clone was derived by Martens *et al.* from a rat-mouse T-cell hybridoma. Toh *et al.* do not mention the possibility that the IgE-BF clone could represent a hybridoma-related recombination event, as opposed to an evolutionarily significant insertion of a retroviral-related sequence into "a cellular DNA region proximal (sic) to a primordial IgE-BF gene". Observations suggesting such an alternative explanation exist in the literature⁴⁻⁶. These data all demonstrate that in the mouse genome, IAP genes are mobile elements which can affect the expression of unrelated genes. Toh *et al.* overlooked the possibility that the structure of the cloned IgE-BF gene could have arisen by insertion (in the hybridoma) of an IAP sequence distal to the IgE-BF coding sequence, followed by transcription of this region and splicing of the "IgE-BF", *gag* and *pol*" sequences. In view of this possibility, it is important to stress that the IgE-BF clone is a complementary DNA, rather than a genomic clone; Toh *et al.* make inferences of considerable import about the structure and evolution of a putative genomic gene from sequence comparisons involving only a cDNA clone.

Taken together, the arguments outlined above also cast doubt on the statement of

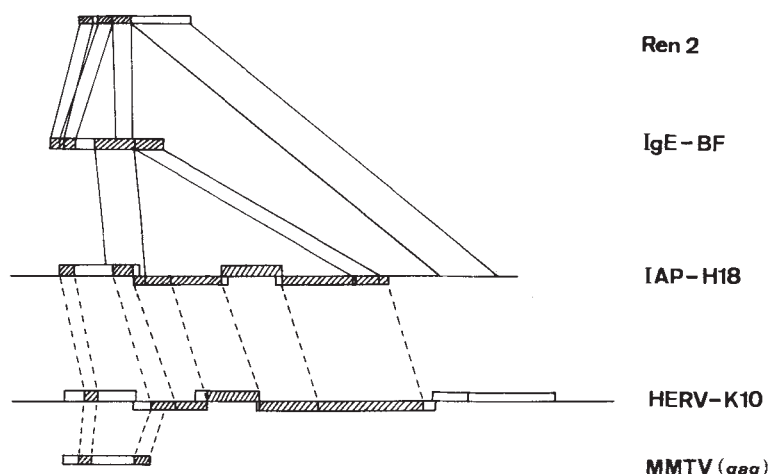


Fig. 1 Profiles of sequence homology of IgE-BF cDNA to mouse (Ren 2) and hamster (H-18) IAPs, human endogenous retrovirus (HERV-K10) and MMTV *gag*. Homologous regions are indicated by slanted lines. Regions where homologies exist at the DNA level and at the protein level are connected by vertical straight lines and dotted lines, respectively.

Toh *et al.* that a particular proteolytic cleavage of the precursor IgE-BF is more likely than another to generate the 11K (relative molecular mass 11,000) IgE-BF from the 60K species.

Thus, we believe that while the observations of Toh *et al.* are of interest, conclusions about their possible significance are speculative, and certainly premature in the absence of any experimental data examining the relationship of IgE-BF to IAP. A preliminary report of the homology of IgE-BF to mouse IAP genes appeared earlier⁷, but was not cited by Toh *et al.*

KEVIN W. MOORE
CHRISTINE L. MARTENS

Department of Immunology,
DNAX Research Institute of
Molecular and Cellular Biology,
Palo Alto,
California 94304, USA

1. Toh, H., Ono, M. & Miyata, T. *Nature* **318**, 388-389 (1985).
2. Martens, C. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2460-2464 (1985).
3. Ono, M., Toh, H., Miyata, T. & Awaya, T. *J. Virol.* **55**, 387-394 (1985).
4. Hawley, R. G., Shulman, M. J., Murialdo, H., Gibson, D. M. & Hozumi, N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7425-7429 (1982).
5. Kuff, E. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 1992-1996 (1983).
6. Kuff, E. L. *et al. Nature* **302**, 547-548 (1985).
7. Martens, C. *et al. in Immune Regulation* (eds Feldmann, M. & Mitchison, N. A.) 121-130 (Humana, Clifton, New Jersey, 1985).

MIYATA ET AL. REPLY—We have shown previously¹ that the IgE-BF cDNA from a rat-mouse T-cell hybridoma² exhibits a strong sequence homology to the Syrian hamster IAP³ at the nucleotide level in the 3' two-thirds of the protein-coding region, whereas the remaining 5' one-third shows no obvious homology. This non-

homologous region of the IgE-BF cDNA, however, shows sequence homology to a mouse IAP (Ren 2)⁴ at the DNA level, but not to a distantly related human endogenous retrovirus, HERV-K10 (M.O. *et al.*, in preparation) or to a mouse mammary tumour virus (MMTV)⁵ *gag* region, both of which share obvious homology with the hamster IAP at the protein level (Fig. 1). Moore *et al.*⁶ recently demonstrated sequence homology of the IgE-BF cDNA to mouse IAPs, represented by MIA14, over its entire regions. These results suggest that the nucleotide sequence corresponding to the 5' third of the IgE-BF coding region is unique to mouse IAPs or their close relatives. This unique region might have been derived very recently from a distantly related retrovirus or an endogenous retrovirus by a recombination mechanism, or from a primordial IgE-BF gene of cellular origin, as suggested previously¹.

TAKASHI MIYATA
HIROYUKI TOH

Department of Biology,
Faculty of Science,
Kyushu University,
Fukuoka 812, Japan

MASAO ONO
Department of Molecular Biology,
School of Medicine,
Kitasato University,
Kanagawa 228, Japan

1. Toh, H., Ono, M. & Miyata, T. *Nature* **318**, 388-389 (1985).
2. Martens, C. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2460-2464 (1985).
3. Ono, M., Toh, H., Miyata, T. & Awaya, T. *J. Virol.* **55**, 387-394 (1985).
4. Burt, D. W., Reith, A. D. & Brammar, W. J. *Nucleic Acids Res.* **12**, 8579-8593 (1984).
5. Fasel, N. *et al. Nucleic Acids Res.* **11**, 6943-6955 (1983).
6. Moore, K. W. *et al. J. Immun.* **136**, 4283-4290 (1986).